

# Integrated Omics and Computational Glycobiology Reveal Structural Basis for Influenza A Virus Glycan Microheterogeneity and Host Interactions\*<sup>§</sup>

Kshitij Khatri‡, Joshua A. Klein‡§, Mitchell R. White¶, Oliver C. Grant||, Nancy Leymarie‡, Robert J. Woods||,  Kevan L. Hartshorn¶, and Joseph Zaias‡\*\*

Despite sustained biomedical research effort, influenza A virus remains an imminent threat to the world population and a major healthcare burden. The challenge in developing vaccines against influenza is the ability of the virus to mutate rapidly in response to selective immune pressure. Hemagglutinin is the predominant surface glycoprotein and the primary determinant of antigenicity, virulence and zoonotic potential. Mutations leading to changes in the number of HA glycosylation sites are often reported. Such genetic sequencing studies predict at best the disruption or creation of sequons for *N*-linked glycosylation; they do not reflect actual phenotypic changes in HA structure. Therefore, combined analysis of glycan micro and macroheterogeneity and bioassays will better define the relationships among glycosylation, viral bioactivity and evolution. We present a study that integrates proteomics, glycomics and glycoproteomics of HA before and after adaptation to innate immune system pressure. We combined this information with glycan array and immune lectin binding data to correlate the phenotypic changes with biological activity. Underprocessed glycoforms predominated at the glycosylation sites found to be involved in viral evolution in response to selection pressures and interactions with innate immune-lectins. To understand the structural basis for site-specific glycan microheterogeneity at these sites, we performed structural modeling and molecular dynamics simulations. We observed that the presence of immature, high-mannose type gly-

cans at a particular site correlated with reduced accessibility to glycan remodeling enzymes. Further, the high mannose glycans at sites implicated in immune lectin recognition were predicted to be capable of forming trimeric interactions with the immune-lectin surfactant protein-D. *Molecular & Cellular Proteomics* 15: 10.1074/mcp.M116.058016, 1895–1912, 2016.

Influenza A virus (IAV)<sup>1</sup> remains a major cause of animal and human mortality and morbidity, with no effective broadly neutralizing vaccines available. The IAV envelope-glycoprotein hemagglutinin (HA) binds sialic acids on respiratory airway surface glycoconjugates and facilitates viral entry into host cells, thereby leading to infection. The envelope-glycoprotein neuraminidase (NA) cleaves sialic acids to allow newly formed virions to escape the host cell membrane. Because the surface proteins undergo rapid mutation, it has not been possible to develop broadly neutralizing vaccines against IAV. In particular, IAV strains undergo amino acid mutations to accumulate new *N*-glycosylation sequons during seasonal circulation in the human population (1–4). Thus, based primarily on genetic sequencing evidence (5–7), glycans appear to shield HA antigenic sites from host antibody recognition, as part of the evolutionary antigenic drift (8–11). Although this putative shielding effect of glycans apparently improves the ability of the virus to escape antibody neutralization, the virus must compensate for loss in receptor avidity from steric interfer-

From the ‡Center for Biomedical Mass Spectrometry, Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118; §Bioinformatics Program, Boston University, Boston, Massachusetts 02215; ¶Department of Medicine, Boston University School of Medicine, Boston, Massachusetts 02118; ||Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602

\* Author's Choice—Final version free via Creative Commons CC-BY license.

Received January 6, 2016, and in revised form, March 4, 2016

Published, MCP Papers in Press, March 16, 2016, DOI 10.1074/mcp.M116.058016

Author contributions: Study design and direction: JZ, KLH. Omics data acquisition: KK, NL. Data analysis: KK, JAK. Bioassays: MRW, KK. Molecular modeling design: RJW. Modeling: OCG.

<sup>1</sup> The abbreviations used are: IAV, Influenza A virus; CAD, Collision-activated dissociation; D+R, NCRD with two amino acid mutations; ERManI, Endoplasmic reticulum class I  $\alpha$ -mannosidase; HA, Hemagglutinin; HCD, Higher-energy collisional dissociation; HILIC, Hydrophilic-interaction liquid chromatography; huNCRD, Wild-type human neck and carbohydrate recognition domain of SP-D; LC-MS: Liquid chromatography-mass spectrometry; MD, Molecular Dynamics; MS/MS, Tandem mass spectrometry; MS<sup>2</sup>, Tandem mass spectrometry; Phil-82, A/Philippines/2/1982 (H3N2); Phil-BS, A/Philippines/2/1982/BS (H3N2) (Bovine serum escape mutant); PR-08, A/Puerto Rico/8/1934 (H1N1); PTM, Post-translational modification; R343V, NCRD with single amino acid mutation; rhSPDII, Recombinant full length human SP-D; SP-D, Surfactant protein-D.

ence caused by the glycans proximal to the sialic acid binding sites on the HA head (12, 13).

T-cells of the adaptive immune system reach the lung approximately 5 days after IAV infection (14). During this time, viral replication and lung damage is mitigated by the innate immune system. Among the soluble innate immune inhibitors, the calcium-dependent lectin inhibitors (known as  $\beta$ -inhibitors) bind to carbohydrates on pathogen surfaces. For seasonal IAV strains, surfactant protein-D (SP-D) is the most important soluble innate immune factor present in bronchial lavage fluid (15, 16). The activity of this lectin depends on the presence of high mannose *N*-glycans on viral HA; IAV strains lacking glycosylation on the HA head region resist neutralization by SP-D. These include the pandemic strains of H1N1 1918 and 2009, the H3N2 of 1968 and the H2N2 of 1957 (17–19). Thus, the pathogenicity of pandemic IAV correlates with ability to escape neutralization by SP-D (15, 18, 20). Generally speaking, as seasonal strains acquire additional *N*-glycosylation sequons they show increased susceptibility to SP-D, as they circulate in humans. Although IAV mutates to avoid antibody neutralization by accumulating sequons, the mechanisms whereby viruses maintain fitness despite increased binding to SP-D and other lectins of the innate immune system remain poorly understood (16, 19, 21–25).

The H3N2 subtype has circulated in humans since 1968 and remains the major seasonal IAV health threat. The H1N1 subtype was introduced into the human population with the disastrous 1918 pandemic and was supplanted by the H2N2 subtype in 1957. The H1N1 subtype was re-introduced into the human population in 1977 and displayed several additional *N*-glycosylation sequons on HA not present on earlier strains (26). In 2009, it underwent antigenic shift by incorporation of genetic elements from the swine and avian sources to emerge as a new pandemic subtype. The pattern of *N*-glycosylation sequons of pandemic H1N1 2009 more closely resembled the 1918 pandemic H1N1 (minimal glycosylation) than that circulating seasonally prior to 2009 (4).

Despite the central role of glycosylation in viral evolution, there remains a paucity of structural information on HA glycosylation. A recent study of engineered variants of the Hong Kong 1968 H3N2 subtype demonstrated that exclusively high-mannose *N*-glycans occupy HA sites 165 and 246 (27). Although genetic sequence analyses identify the creation or disruption of *N*-glycosylation sequons, they provide no information on the glycan site occupancy, structure and microheterogeneity. In addition, crystal structures typically show only the chitobiose core and do not define glycan antennae (28–30). We therefore built a predictive model for seasonal IAV fitness and antigenicity that incorporates (1) data on changes in site-specific HA glycosylation that occur as the virus evolves to escape SP-D neutralization and (2) the HA structural features that regulate glycan microheterogeneity at each site that influences binding to SP-D. We demonstrate the value of complete information on the mature glycosylated

HA structures to inform better understanding of the evolution of IAV in response to immune system pressure.

#### EXPERIMENTAL PROCEDURES

**Virus Preparation**—The following IAV strains were used in the study:

**Phil-82**—A high growth reassortant of A/Philippines/2/82(H3N2) and A/Puerto Rico/8/34(H1N1). The seed stock was a gift from Dr. E. Margot Anders, University of Melbourne. This strain consisted of an H3 hemagglutinin and N2 neuraminidase from A/Philippines/2/82, whereas the remaining virus components were from A/Puerto Rico/8/34.

**Phil-BS**—A mutant version of Phil-82 that was selected by growing Phil-82 in the presence of bovine serum, found to be resistant to neutralization by bovine conglutinin, as described previously (22–24, 31). The seed stock was from Dr. E. Margot Anders, University of Melbourne.

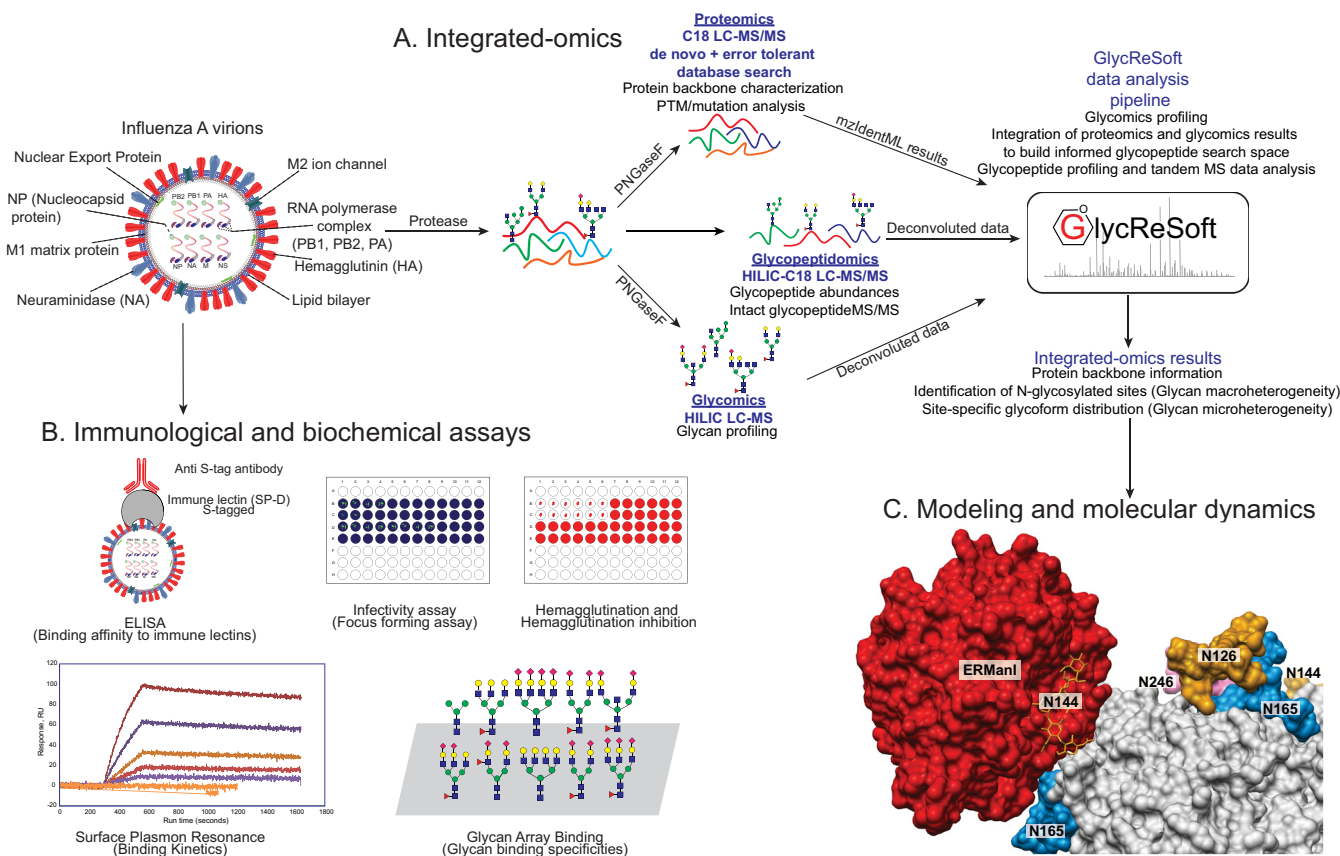
**PR-08**—The mouse-adapted A/Puerto Rico/8/34(H1N1) strain. The seed stock was from Dr. Jon S. Abramson, Wake Forest University.

All viruses were grown in the chorioallantoic cavity of 10 day old embryonated chicken eggs. Allantoic fluid was harvested after incubating for 48 h and insoluble debris were pelleted by centrifugation at  $1000 \times g$ . The virions were then pelleted at  $135,000 \times g$  and the pellets were purified on a sucrose density gradient followed by dialysis against phosphate buffered saline (PBS).

**Experimental Design and Statistical Rationale**—As shown in Fig. 1, virus samples were subjected to proteolytic digestion. Proteolytic samples were divided into three fractions for proteomics, glycomics and glycoproteomics. Proteomics and glycomics fractions were subjected to PNGase *F* glycan release. All three fractions for each sample were subjected to triplicate LC-MS analyses, as described in sections below. Three technical replicates of LC-MS acquisitions were used to calculate mean analyte abundances and standard deviations in measurements. Human  $\alpha$ -1-acid glycoprotein (Sigma-Aldrich, St. Louis, MO), a widely studied standard glycoprotein, was used as a control for all sample preparation and verifying LC-MS instrument performance.

#### LC-MS Analyses

**Sample Preparation**—Virus preparations were subjected to membrane disruption, reduction, alkylation and proteolysis. Virus preparations containing  $\sim 60 \mu\text{g}$  total protein were dried down in a centrifugal evaporator and re-suspended in LC-MS grade methanol (Fisher Scientific), followed by sonication for 10 min in a water bath to disrupt viral membranes. Samples were dried and re-suspended in 50% 100 mM ammonium bicarbonate and 50% 2, 2, 2-trifluoro ethanol (TFE) (Sigma-Aldrich) ( $60 \mu\text{l}$ ). 200 mM dithiothreitol (DTT) ( $2 \mu\text{l}$ ) was added to the samples and heated at  $95^\circ\text{C}$  for 30 mins to reduce disulfide bonds. Iodoacetamide ( $10 \mu\text{l}$ , 200 mM in water) was then added to the samples and incubated for 1 h in dark, at room temperature. To quench excess iodoacetamide, another ( $2 \mu\text{l}$ ) of 200 mM DTT was added to the samples and incubated for one hour at room temperature. LC-MS water ( $300 \mu\text{l}$ ) was added to the samples to dilute TFE to less than 5% of the total volume and 100 mM ammonium bicarbonate ( $200 \mu\text{l}$ ) was added to raise pH to 7.5. Sequencing grade trypsin or chymotrypsin (Promega Corp., Madison, WI) were added to the samples and incubated overnight at  $37^\circ\text{C}$ , for proteolysis. Proteolytic peptides and glycopeptides were split into aliquots for performing  $^{16}\text{O}$ -proteomics,  $^{18}\text{O}$ -proteomics (for site-occupancy analysis) and glycoproteomics.  $^{16}\text{O}$ -proteomics and  $^{18}\text{O}$ -proteomics samples were deglycosylated using PNGase *F* (New England Biolabs, Ipswich, MA) in  $\text{H}_2^{16}\text{O}$  or  $\text{H}_2^{18}\text{O}$  respectively, whereas the glycoproteomics samples were incubated in the deglycosylation conditions without PNGase *F*. Released glycans from deglycosylated samples were isolated using C18 spin columns (Pierce Biotechnology, Rockford, IL) and



**FIG. 1. Experimental Workflow.** *A*, Integrated-omics: Workflow for acquiring and combining proteomics, glycomics and glycopeptidomics information to enable confident assignment of site-specific glycoforms; *B*, Immunological and biochemical assays used for correlating changes in virus glycosylation with bioactivity; *C*, Modeling and molecular dynamics simulations driven by structural information from integrated-omics analyses to understand structural basis for host-virus interactions and glycan processing at important sites. Graphic shows ERMAnI interacting with the glycan at Asn<sup>144</sup> on the head region of a Phil-82 hemagglutinin trimer.

pooled (<sup>16</sup>O and <sup>18</sup>O) for glycomics analyses, and deglycosylated peptides were eluted and collected separately for proteomics analyses. As a control for spontaneous deamidation at inherently nonglycosylated asparagine residues, the glycosylated samples (no PNGase *F* treatment) were also analyzed, using the proteomics workflow. Mass spectrometry data were acquired on each sample as three technical replicates, unless otherwise stated.

**Proteomics Analyses**—Deglycosylated peptides were subjected to LC-MS/MS using a Waters<sup>TM</sup> NanoAcquity<sup>TM</sup> nano-flow chromatograph (Waters Corp., Milford, MA) mounted with a Waters<sup>TM</sup> Xbridge<sup>TM</sup> reversed-phase column (150  $\mu$ m  $\times$  100 mm) packed with 1.7  $\mu$ m BEH C18 resin and a Waters<sup>TM</sup> trap column (180  $\mu$ m  $\times$  20 mm) packed with 5  $\mu$ m Symmetry<sup>TM</sup> C18 stationary phase. The chromatography was performed online with a Q-Exactive Plus<sup>TM</sup> mass spectrometer (Thermo Fisher Scientific<sup>TM</sup>, San Jose, CA), operated in positive ion mode. The mass spectrometer was mounted with an Advion NanoMate (Advion Inc., Ithaca, NY) source for introduction of the LC eluate by nano-ESI. The source was operated at 1.7 kV with transfer-capillary temperature maintained at 250  $^{\circ}$ C and S-Lens RF level set at 55. Mass spectra were acquired in the Orbitrap mass analyzer with 1 microscan per spectrum for both MS and MS/MS. Resolving power for MS and MS/MS were set at 70,000 and 17,500, respectively. Tandem MS data were acquired in parallel with MS, on the top 20 most abundant multiply charged precursors, with higher energy collisional dissociation (HCD) at normalized collision energy of

27V. Precursors were isolated using a 1.4 Th window and dynamic exclusion of 10 s was enabled during precursor selection.

**Glycomics Analyses**—Released *N*-glycans were desalted using a Superdex<sup>TM</sup> Peptide<sup>TM</sup> 3.2/30 size exclusion column on a Beckman<sup>TM</sup> Gold chromatograph. Desalted *N*-glycans were analyzed in triplicate using negative mode hydrophilic interaction liquid chromatography (HILIC) -MS with a constant acetonitrile post-column make-up-flow on an Agilent<sup>TM</sup> 6520 LC-MS system with chip-cube<sup>TM</sup> nanoESI source, as described previously (32).

**Glycoproteomics Analyses**—Entire tryptic digests were subjected to triplicate analyses using chip-LC-MS/MS with online HILIC enrichment and C18 separation on an Agilent<sup>TM</sup> 6550 LC-MS system with chip-cube<sup>TM</sup> ionization source as described previously (33) and by C18 LC-MS/MS without any enrichment on the Q-Exactive Plus<sup>TM</sup> instrument as described for the proteomics samples. Chymotryptic samples were analyzed only using C18 LC-MS/MS on the Q-Exactive Plus<sup>TM</sup>.

#### Data Analysis—

**Proteomics**—Proteomics data were subjected to PEAKS SPIDER (*de novo* + database + error tolerant PTM) searches in PEAKS (34) Studio 7.5 (Bioinformatics Solutions Inc., Waterloo, ON) against a combined Uniprot (35) and Influenza A virus protein database (36). The FASTA sequence database used for searches has been included in the data repository. A 50 ppm error tolerance for the precursor (MS<sup>1</sup>) and 0.1 Da mass error tolerance for fragment ions (MS<sup>2</sup>) were

specified. A maximum of 2 missed cleavages per peptide were allowed for the database search. Trypsin specificity was defined as cleavage after Arg and Lys, when not followed by a Pro. Chymotrypsin cleavage specificity was after Phe, Leu, Met, Trp and Tyr residues, when not followed by a Pro. Cysteine carbamidomethylation was specified as a fixed modification; deamidation (Asn) and oxidation (Met) were specified as variable modifications.  $^{18}\text{O}$  labeling of deglycosylated asparagine was also used as a variable modification in case of  $^{18}\text{O}$ -proteomics samples. After a regular database search, an error-tolerant PTM search was also performed searching for a larger subset of modifications from Unimod (37) and any amino acid substitutions. The final results were a combination of database, *de novo* and error-tolerant searches. False discovery rates (FDR) were calculated using a decoy-fusion approach in PEAKS 7.5, as described previously (34). Identified peptide-spectrum-matches with  $-10\log\text{P}$  value of 15 or higher were kept, at a FDR threshold of 0.1%.

**Site-occupancy Analysis**—Site-occupancy analysis was performed on Phil-82 for glycan modeling and molecular dynamics. Although  $^{18}\text{O}$ -proteomics was performed to define glycan site-occupancy, in cases where multiple asparagine and glutamine residues existed on peptides containing the *N*-glycosylation sequon, spontaneous deamidation of these residues in presence of  $\text{H}_2^{18}\text{O}$ , confounded the data. Therefore,  $^{16}\text{O}$ -proteomics results were instead used to define site-occupancy information.

From the proteomics search results, peptides spanning glycosylation sites were selected and extracted ion chromatograms were integrated to generate chromatographic peak areas for paired analyses of unoccupied peptides bearing glycosylation sequon in glycosylated samples and the same peptides in PNGaseF deglycosylated sample. Average peak areas of three replicate analyses were compared after normalization using a high-confidence peptide from a different protein identified across runs. Site-occupancy for sites 8 and 22 were calculated from proteomics results of chymotryptic digests, whereas for the remaining sites, proteomics results from tryptic samples were used.

Percent site occupancy was calculated as

$$\text{Site occupancy} = \frac{\text{Abundance (deglycosylated)} - \text{Abundance (glycosylated)}}{\text{Abundance (deglycosylated)}}$$

(Eq. 1)

**Glycomics**—Negative mode HILIC-MS data were deconvoluted and deisotoped using the THRASH algorithm in DeconTools/Decon2LS (version 1.0.5501) (38, 39). The deconvoluted/deisotoped peaklists were matched against a theoretical composition hypothesis containing human and avian *N*-linked glycans from GlycomeDB (40, 41), using GlycReSoft (42). Up to 2 formate adducts were allowed while searching for glycan compositions, with a match error tolerance of 20 ppm in the theoretical masses. Abundances of different adducted forms were grouped together and reported as mean of three technical replicates.

**Glycoproteomics/Integrated-omics**—Glycopeptide LC-MS/MS data were deconvoluted and deisotoped using DeconTools (39) for MS<sup>1</sup> spectra. Data were converted to mzML format (43) using the MSConvert tool in Proteowizard (version 3.0.7692) (44) and the MS<sup>2</sup> spectra were then deconvoluted using MasSpike (45). Deconvoluted data were analyzed using an in-house version of GlycReSoft and validated manually. Source code for GlycReSoft is publically hosted on GitHub (<https://github.com/GlycReSoft2>). Naïve and informed glycopeptide hypotheses were created for mining glycopeptide data. For naïve hypothesis generation, theoretical digests of glycoproteins were generated for combining *N*-glycosylation sequon containing peptides with *N*-glycan compositions from GlycomeDB (41) glycan databases.

For informed hypothesis generation, proteomics search results exported in mzIdentML 1.1 format, were used. Hemagglutinin pep-

tides that contained Asn from the *N*-linked glycosylation sequon, with  $-10\log\text{p}$  values of 20 or higher were included in the hypotheses. Peptide variants were combined with glycoforms identified in glycomics data, to build the glycoproteomics hypotheses. MS<sup>1</sup> deconvoluted data were searched against glycopeptide hypotheses, using a match error-tolerance of 20 ppm. Theoretical fragment ions were generated for the MS<sup>1</sup> matches and then searched against the MS<sup>2</sup> deconvoluted peak lists, using a 20 ppm error tolerance. Features searched for in MS<sup>2</sup> data of glycopeptides included oxonium ions, peptide backbone ions, peptide backbone ions with an attached HexNAc and stub-glycopeptide ions (33). A glycopeptide spectrum match, was required to have mono or disaccharide oxonium ions. Automated glycopeptide analysis results have been included in the data repository. MS<sup>1</sup> scores were calculated as described previously (42). MS<sup>2</sup> scores are representative of number of fragment ions found (peptide backbone and stub ion coverage) (33). *q*-values depict minimal FDR threshold at which the identification is accepted, as described by Käll *et al.* (46). For cases, where more than one glycopeptide composition matched a given precursor and fragment ion spectrum, only the highest scoring match was retained for reporting in the results. Glycopeptide analysis results represent site-specific glycoform abundances for glycopeptides assigned confidently based on MS<sup>2</sup> spectra by GlycReSoft and validated manually. Chymotryptic glycopeptide data were manually analyzed for resolving glycoforms on Asn<sup>8</sup> and Asn<sup>22</sup> on Phil-82 and Phil-BS and integrated with site-specific glycosylation data from tryptic samples processed using GlycReSoft.

**Bioassays**—Biochemical and immunological assays were performed to correlate changes in site-specific glycan distributions in IAV strains with virus fitness and susceptibility toward host innate immune system. SP-D is a collectin produced and secreted by lung epithelial cells in humans and other mammals. SP-D is also known to inhibit IAV by binding its surface glycoproteins HA and NA (15, 31, 20, 47). For this study, a recombinant full length human SP-D (rhSPDII) that exists as dodecamers was prepared, as described previously by Nikolaidis *et al.* (48, 49). In addition to the full-length SP-D, recombinant neck and carbohydrate recognition domain (NCRD) from SP-D were also used. These NCRDs lack the *N*-terminal domain and the extended collagen domain present in the full-length SP-D and as a result form homotrimers instead of higher order multimers. The wild-type recombinant human NCRD (huNCRD) has been shown to bind glycans but lacks antiviral activity because of absence of cooperative binding effect from the multiple heads (49–51). Gain-of-function mutants R343V and double mutant D325A+R343V (D+R), which have been shown to possess increased virus binding and neutralizing activity (48, 49, 52–54), were also used in bioassays. Mutant names are indicative of amino acid substitutions flanking the lectin site in the huNCRD protein.

**Hemagglutination Assays**—Hemagglutination and hemagglutination inhibition (HI) assays were performed in round bottom 96 well plates, using PBS containing calcium and magnesium (PBS+ +) as a diluent. Viral titers were first measured using hemagglutination assays, performed as described previously (55), using type O- red blood cells from human donors. HI assays were performed with recombinant SP-D preparations described above. Inhibition or disruption of agglutination was measured using serial dilutions of SP-D preparations.

**Enzyme-linked Immunosorbent Assay (ELISA)**—96-well plates were coated overnight at 4 °C with 0.1 μg/well of Phil-82, Phil-BS and PR-08. Serial dilutions of NCRDs were added to control wells. Wild-type huNCRD and mutant version D+R were used. Plates were washed 3x with 100 μl of PBS++ (PBS containing 0.9 mM CaCl<sub>2</sub>, 0.52 mM MgCl<sub>2</sub> and 0.16 mM MgSO<sub>4</sub>) and blocked with blocking buffer (Superblock, Pierce). To the wells coated with IAV, different doses of NCRD, matching those in control wells, were added and incubated for

30 min at 37 °C. Plates were then washed 3× with wash buffer. To each well, horseradish peroxidase conjugated anti-S-protein antibody was added and incubated at 37 °C. Following incubation, plates were washed 3× with PBS++ and 50  $\mu$ l of 1-step ELISA substrate (Pierce) was added. After 20 min incubation at room temperature with the ELISA substrate, the reaction was stopped by addition of 50  $\mu$ l 1N sulfuric acid. Plates were read for absorption at 450 nm wavelength with wavelength correction at 540 nm.

**Fluorescent Focus assay (IAV infectivity assays)**—Infectivity assays were performed using human bronchial/tracheal epithelial (BTE) cells (Lifeline Cell Technology, Walkersville, MD). BTE cell monolayers were grown to about 70% confluency in 96 well plates. IAV preparations were pre-incubated with SP-D preparations for 30 min at 37 °C, to allow SP-D binding and inhibition of HA. The BTE cell monolayers were infected with these IAV preparations for 45 min at 37 °C, in PBS++. Cells were tested for presence of IAV after 18 h of virus addition using an anti IAV nucleoprotein monoclonal antibody (Millipore) and fluorescence detection, as described previously (56).

**Surface Plasmon Resonance**—Surface Plasmon Resonance (SPR) experiments were performed as described previously by Crouch *et al.* (53). Recombinant human SP-D NCRD (Neck and carbohydrate recognition domain) preparations were immobilized using N-terminal His tags on NiNTA, with the binding domain oriented upward. Binding of Phil/82 and Phil/82/BS B-HA were measured against recombinant wild type huNCRD and double mutant D+R.

**Glycan Array Binding**—Intact virions were labeled with Alexa 488 succinimidyl ester conjugated dye (Life Technologies/Thermo Fisher Scientific, Carlsbad, CA) using the manufacturer's protocol. Postlabeling, virions were dialyzed against 1× PBS. HA titers were measured and the virus preparations were appropriately diluted to adjust for differences in HA titers before binding experiments. Glycan array binding was performed using standard protocols at the Consortium for Functional Glycomics (57–59) Protein-Glycan Interaction Core at Emory University, using array version 5.2 (<https://glycopattern.emory.edu/structures/view?version=5.2>). Binding was measured as fluorescent units of bound fluorescently labeled influenza virions.

#### Computational Glycobiology—

**3D Structure Generation**—A homology model for the Phil/2/1982 HA trimer was generated in Modeler (60–62) using PDBIDs 2YP3, 2YP7, 1HA0, 4KVN, 4WE5, 4O58, and 4WA1 as templates. In order to decrease the time required to perform a molecular dynamics (MD) simulation, a smaller structure was generated from the HA0 trimer that contained only residues F83 to D86 of the stalk and D57 to S270 of the head group. Glycans were attached to the head group using a customized version of the glycoprotein builder available on GLYCAM-Web (63), with the following modifications. When attaching the N-glycan to an Asn sidechain, the  $\chi_1$ ,  $\chi_2$ ,  $\psi$ , and  $\phi$  torsion angles were adjusted to match observed rotamers reported for N-linked glycoproteins in the PDB (64). Any vdW overlaps were relieved by adjusting the glycosidic torsion angles within normal bounds (65). In all cases, the Asn sidechain conformation that was observed to be most common in the PDB was selected, provided that it could be attached without irreconcilable steric overlaps with the protein surface. Man<sub>9</sub>GlcNAc<sub>2</sub> glycans were added to each of the five N-glycosylation sites on the HA headgroup; at sites Asn<sup>63</sup>, Asn<sup>126</sup>, Asn<sup>144</sup>, Asn<sup>165</sup>, and Asn<sup>246</sup>.

**System Preparation for Simulation**—hydrogen atoms were added to the head group structure, with the protonation states of the ionizable side chains being assigned by tleap (66). The structures were placed in a periodic box of ~35,000 TIP5P waters (67) to provide an 8 Å buffer between the glycan and the edge of the periodic box.

**Energy Minimization and Solvent Equilibration**—energy minimization of all atoms was performed for 20,000 steps (10,000 steepest decent, followed by 10,000 conjugate gradient). The energy-minimized structures were equilibrated at 300 K under nPT conditions for

400 ps, with 5 kcal/mol-Å<sup>2</sup> Cartesian restraints on the solute heavy atoms.

**System Equilibration and Production MD**—after solvent equilibration, the heavy atom restraints were removed, and 1 ns of equilibration MD followed by a 500 ns production simulation were performed. The C $\alpha$  atoms of the terminus residues were restrained (5 kcal/mol-Å<sup>2</sup>).

**Energy Minimization and Molecular Dynamics (MD) Simulations**—all simulations were performed with the CUDA implementation of PMEMD (68, 69) in the Amber14 software suite (66). The GLYCAM06-j (63) and Amber14SB parameters (66) were employed for the carbohydrate and protein portions, respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, whereas a Langevin thermostat with a collision frequency of 2 ps<sup>-1</sup> was employed for temperature regulation. A nonbonded interaction cutoff of 8 Å was employed. Long-range electrostatics were treated with the particle-mesh Ewald (PME) method (70). Covalent bonds involving hydrogen were constrained with the SHAKE algorithm allowing a time step of 2 fs (71).

**Post Processing**—An ensemble of 1000 conformations for the glycosylated Phil-2-1982 HA glycoprotein were taken at regular 500 ps intervals of the MD simulation using cpptraj (72).

**Assessment of Glycan Accessibility to ERMAnI**—A new crystal structure of ERMAnI, kindly provided by Dr. Kelley Moremen's laboratory, includes a full high mannose glycan in the binding site. This allowed assessment of the accessibility of the glycan to ERMAnI by superimposing the ERMAnI-bound Man $\beta$ 1-4GlcNAc $\beta$  residues onto each N-glycan on the H3 head group. This was performed for each N-glycan over 1000 snapshots taken at regular intervals from the MD simulation. If surface overlaps greater than ~36 Å<sup>2</sup> (equivalent to a buried Carbon atom) were observed, the glycan was deemed to be inaccessible for that particular time point of the simulation.

**Assessment of Glycan Accessibility to SP-D**—A 3D structure has been determined previously for unliganded human SP-D (PDBID 1B08), as well as porcine SP-D in complex with Man $\alpha$  (PDBID 4DN8) (73). Despite structural differences in the carbohydrate binding sites of the porcine and human proteins, the relative spacing of the binding sites is identical, with a linear distance of ~45 Å between each Man $\alpha$  binding site. By monitoring the distances between the non-reducing terminal Man $\alpha$  residues in each N-glycan over the course of the simulation, it was determined which sets of three N-glycans were positioned for trimeric interaction with SP-D (sets that were all 43–47 Å apart for any of the 1000 snapshots).

## RESULTS AND DISCUSSION

We modeled the changes in HA mature structure that occur as an H3N2 subtype evolves during repeated growth in eggs in the presence of bovine serum. This process rendered the virus resistant to conglutinin and mannose binding lectin in serum, and to SP-D as well (15). Owing to the error prone nature of virus replication and the inclusion of host proteins in the virion structure (74), we acquired deep proteomes of H3N2 before and after development of resistance to SP-D. We then profiled the range of glycans present on the virions and used this information to drive determination of confident site-specific glycosylation from glycoproteomics data. For HA, the combination of peptide proteolytic variants, other PTMs including oxidation and deamidation, and glycosylation yield an extremely large number of theoretical structures. To address this, we combined proteomics, glycomics and glyco-

proteomics analyses to produce site-specific glycosylation assignments.

We and others have developed liquid-chromatography-mass spectrometry methods with sufficient power for analysis of glycopeptides from complex glycoproteins (33, 75–78). In addition, improvements to software tools further enhance the ability to analyze complex glycoproteins (79, 80). We built an exhaustive glycopeptide bioinformatics search-space that made no assumptions regarding peptide backbone, proteolytic variants, PTMs, or glycome present in the IAV samples. We then defined the mature structures of H3N2 HA and the changes associated with escape from SP-D neutralization. We used viral infectivity assays, glycan array binding, lectin binding and other biochemical methods to demonstrate the changes in biological interactions and activities associated with evolution of glycosylation structure among IAV subtypes and strains. In order to understand how particular glycoforms are generated, we modeled the interactions between mature glycosylated HA and ER class I  $\alpha$ -mannosidase (ERManI), a key enzyme in the glycan biosynthetic pathway. We then defined the key glycans mediating interactions of SP-D with HA using structure modeling and molecular simulation studies.

**Proteomics**—Influenza A shows host species-specificity in infection governed by HA binding to sialylated glycans present on host cell-surface receptors. Avian-strains prefer binding to NeuAc- $\alpha$ -2,3-Gal, whereas human-strains prefer binding to NeuAc- $\alpha$ -2,6-Gal (81–84). Minor changes to the protein sequence correlate with a switch in sialic-acid binding preferences leading to changes in receptor specificities and altered susceptibility toward antibodies and lectins (17, 85–87).

We performed deep proteomics on deglycosylated IAV digests. Despite the fact that IAV contains only 8 gene segments, corresponding to roughly 11 proteins, the virions include host proteins as integral parts of its architecture (74). In addition, host proteins may copurify with the virions. We therefore accounted for all proteins and glycoproteins present in the samples in order to maximize the confidence of site-specific glycosylation assignments. The error-prone manner of IAV replication allows incorporation of mutations in viral proteins that may affect glycosylation sites and neighboring residues. Our deep proteomics analyses of virus preparations covered all glycosylation sites and identified glycopeptide variants that included other PTMs and incomplete proteolytic cleavages. We therefore used an unbiased bioinformatics search space for glycopeptide assignment.

In addition to identifying peptide backbone variants, we identified a revertant population in the Phil-BS sample. The *N*-glycosylation sequon at Asn<sup>165</sup> apparently became disrupted because of an acquired mutation leading to conversion of NVTM to NVAM in the majority of Phil-BS virion population; however, a small fraction of the population con-

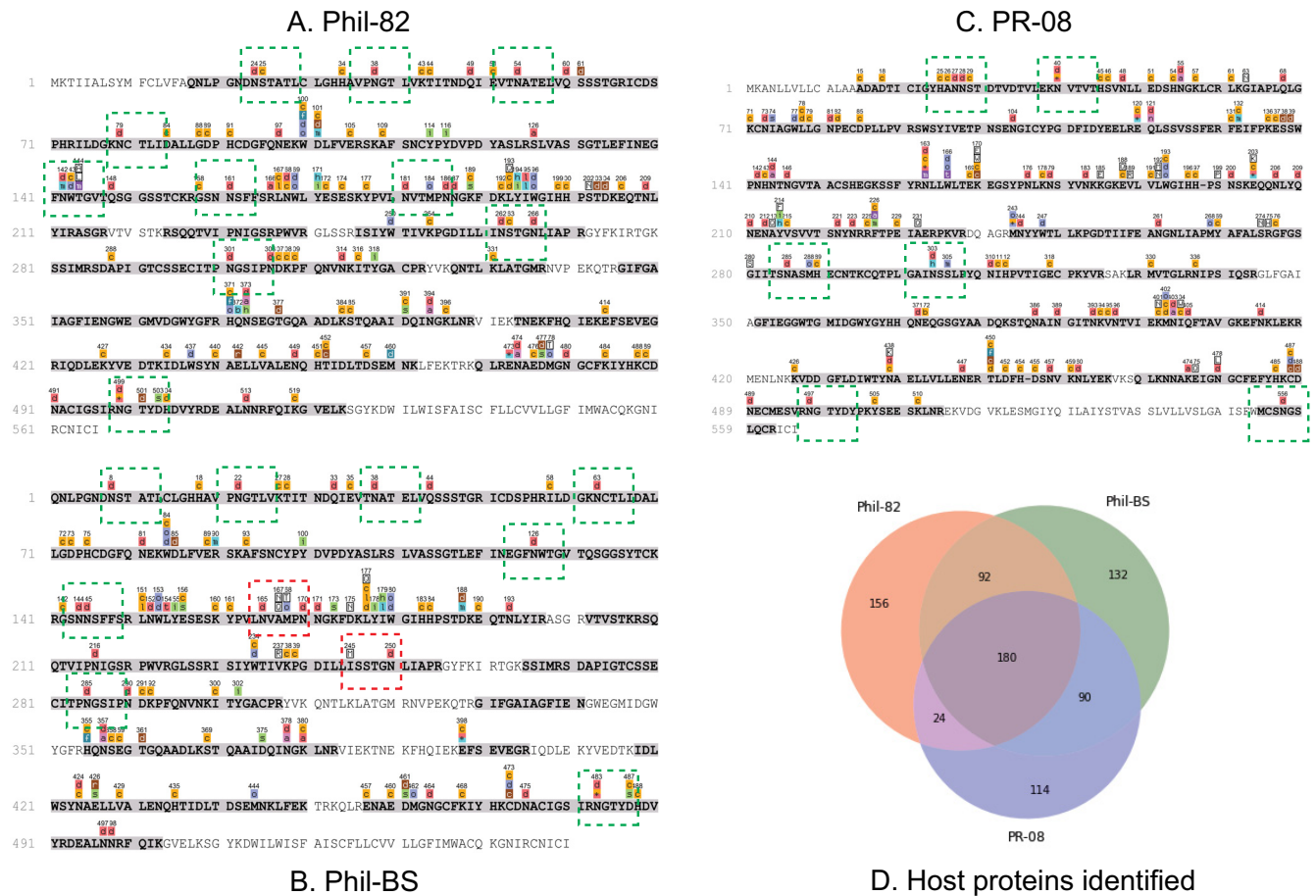
tained the wild-type sequon NVTM. Based on extracted ion chromatogram peak areas, the revertant corresponded to ~2% of the population (see Site occupancy analysis data). We observed no reversion at the second disrupted sequon (Asn<sup>246</sup> NSTG) mutated in Phil-BS. To check for presence of glycosylation at the revertant sequon, we included both Phil-82 and Phil-BS protein sequences in the Phil-BS search hypothesis.

A majority of proteins identified were common among the three IAV samples (Fig. 2D). Gene ontology analysis performed using STRAP (88), showed similar distributions of proteins in terms of molecular functions and cellular localization (see results in the Functional annotation of proteomics results). Functional annotation of the proteins showed close similarity to the results described by Hutchison *et al.* (74), showing a wide-variety of host-proteins involved in various structural and molecular roles present in the virions. We observed deamidation (NQ) as the most commonly encountered PTM in the proteomics results, followed by nonspecific carbamidomethylation, oxidation (M), sodium adduction and dehydration (supplemental Fig. S2).

We identified 10 and 8 sequons for Phil-82 and Phil-BS respectively, as shown in Fig. 2 and supplemental Fig. S1. All the peptide variants identified and used to construct our informed glycoproteomics bioinformatics hypotheses are shown in Protein and peptide sequences used in proteomics and glycoproteomics searches.

**Glycan site-occupancy analysis**—For modeling and molecular-dynamics studies, we determined glycosylation site occupancy of Phil-82 HA. We evaluated the NetNGlyc server (89) (<http://www.cbs.dtu.dk/services/NetNGlyc/>), used in published studies (90–93, 27) to provide a qualitative measure of the likelihood for *N*-linked glycosylation, given the protein sequence; however, we found that NetNGlyc predictions inconsistent with our results. Table I compares NetNGlyc predictions and site-occupancy results from proteomics data for Phil-82. Although NetNGlyc predicted no glycosylation at sites 38 and 144, we identified glycopeptides containing those two sites with high confidence from glycopeptide tandem MS data. The data showed low site occupancy at Asn<sup>144</sup> and high occupancy for Asn<sup>38</sup>. These results re-emphasize the importance of analyzing both deglycosylated peptides and intact glycopeptides to minimize assumptions about samples and shifting reliance from prediction tools to empirical evidence.

**Glycomics**—We used glycomics profiling to define the range of glycans present on virions precisely. The glycan profiling results for Phil-BS (38% high-mannose) showed an ~10% decrease in relative abundances of high-mannose type *N*-glycans compared with Phil-82 (48%) (Fig. 3A). PR-08 displayed a very small proportion of *N*-glycans (ca. 6%) occurring as the high mannose type. Detailed comparison of abundances for individual glycan compositions (Fig. 3B, 3C, 3D) showed a marked decrease in the relative abundances of



**Fig. 2. Proteomics results.** Proteomics coverage of Phil-82 (A), Phil-BS (B) and PR-08 (C) hemagglutinins. Portions of sequence covered are highlighted in gray, with colored symbols above the sequence rows representing modifications and mutations. Green boxes show *N*-glycosylation sequons and red boxes indicate mutated/disrupted sequons in Phil-BS; D: Total number of proteins including host-proteins identified in proteomics of the three IAV samples. Only IAV and chicken (*Gallus gallus*) proteins identified with 2 or more unique peptides were counted.

TABLE I

*NetNGlyc* predictions for glycosylation site occupancy and site-occupancy from proteomics data for Phil-82

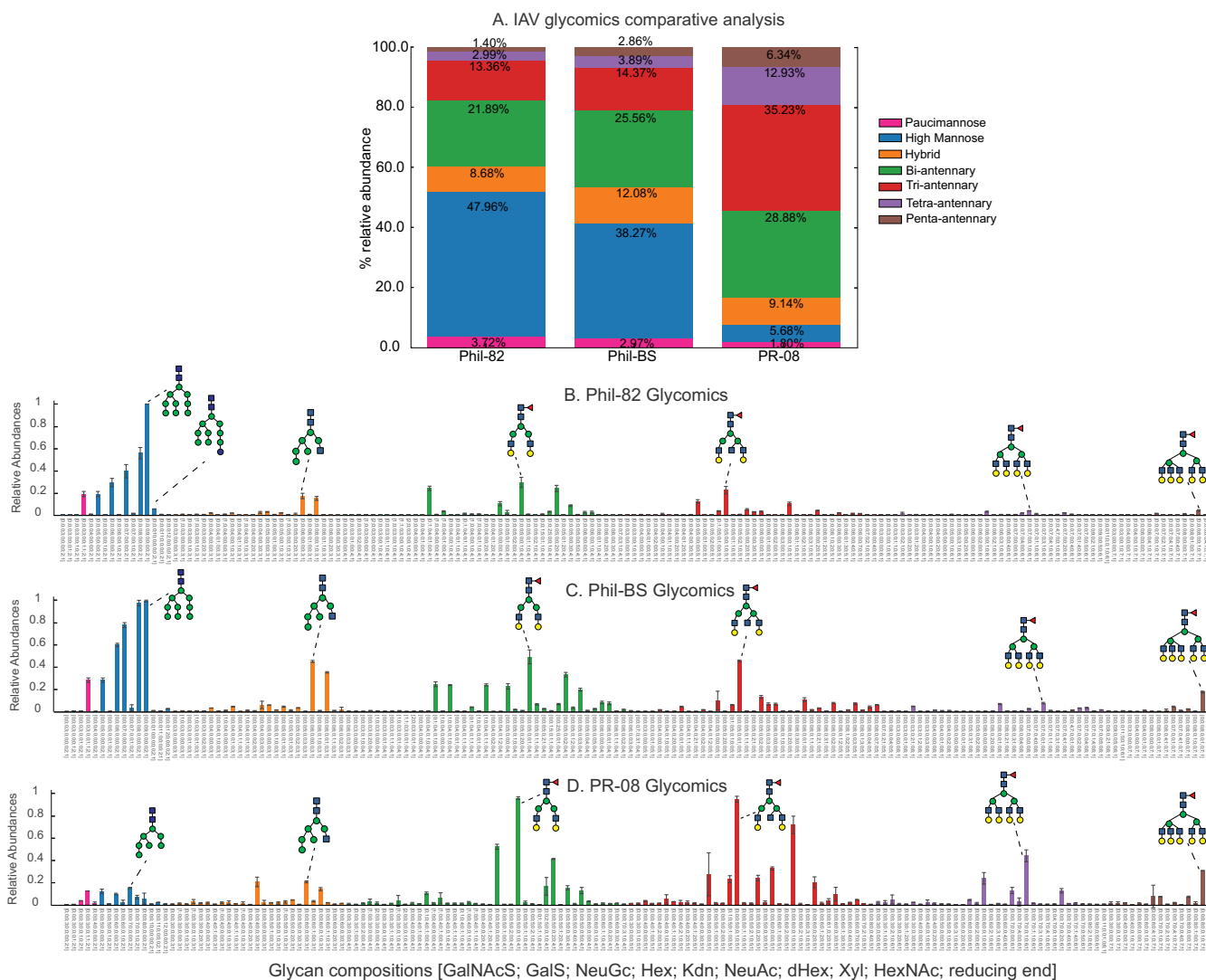
Site	NetNGlyc Potential	NetNGlyc Jury Agreement	NetNGlyc Results	Site-occupancy calculated from proteomics
8 NSTA	0.7987	9/9	+++	1.0
22 NGTL	0.7153	9/9	++	1.0
38 NATE	0.4996	3/9	-	0.99
63 NCTL	0.6366	9/9	++	0.99
126 NWTG	0.5393	6/9	+	0.73
144 NNSF	0.4293	6/9	-	0.17
165 NVTM	0.7626	9/9	+++	0.99
246 NSTG	0.5880	7/9	+	0.99
285 NGSY	0.6716	9/9	++	0.95
483 NGTY	0.5207	8/9	+	0.95

immature Hex<sub>9</sub>HexNAc<sub>2</sub> and Hex<sub>10</sub>HexNAc<sub>2</sub> glycan compositions in Phil-BS. The glycan topologies shown in Fig. 3 are speculative. Glycomics profiling using LC-MS provides saccharide compositions only—topological/structural information cannot be derived from these experiments. Detailed re-

sults from glycomics analyses have been included in the data repository.

From the proteomics results, the primary difference between Phil-82 and Phil-BS was the disruption of two *N*-linked glycosylation sequons at positions 165 and 246. The glycomics results implied that a large proportion of the glycans occupying these two sequons were high-mannose type. To confirm this theory, we carried out intact glycopeptide analysis.

**Glycoproteomics**—For confident assignment of site-specific glycoforms, we analyzed intact glycopeptides using LC-tandem-MS and mined the data using both naive and informed bioinformatics search space hypotheses, as described in the experimental section. Collisional dissociation of glycopeptides in the gas-phase generates abundant ions from glycosidic bond dissociation; however, provided application of sufficient collision energy, peptide bond dissociation also occurs (33). Fig. 4 compares the features in collisional dissociation of glycopeptides using low *versus* high collision energies.



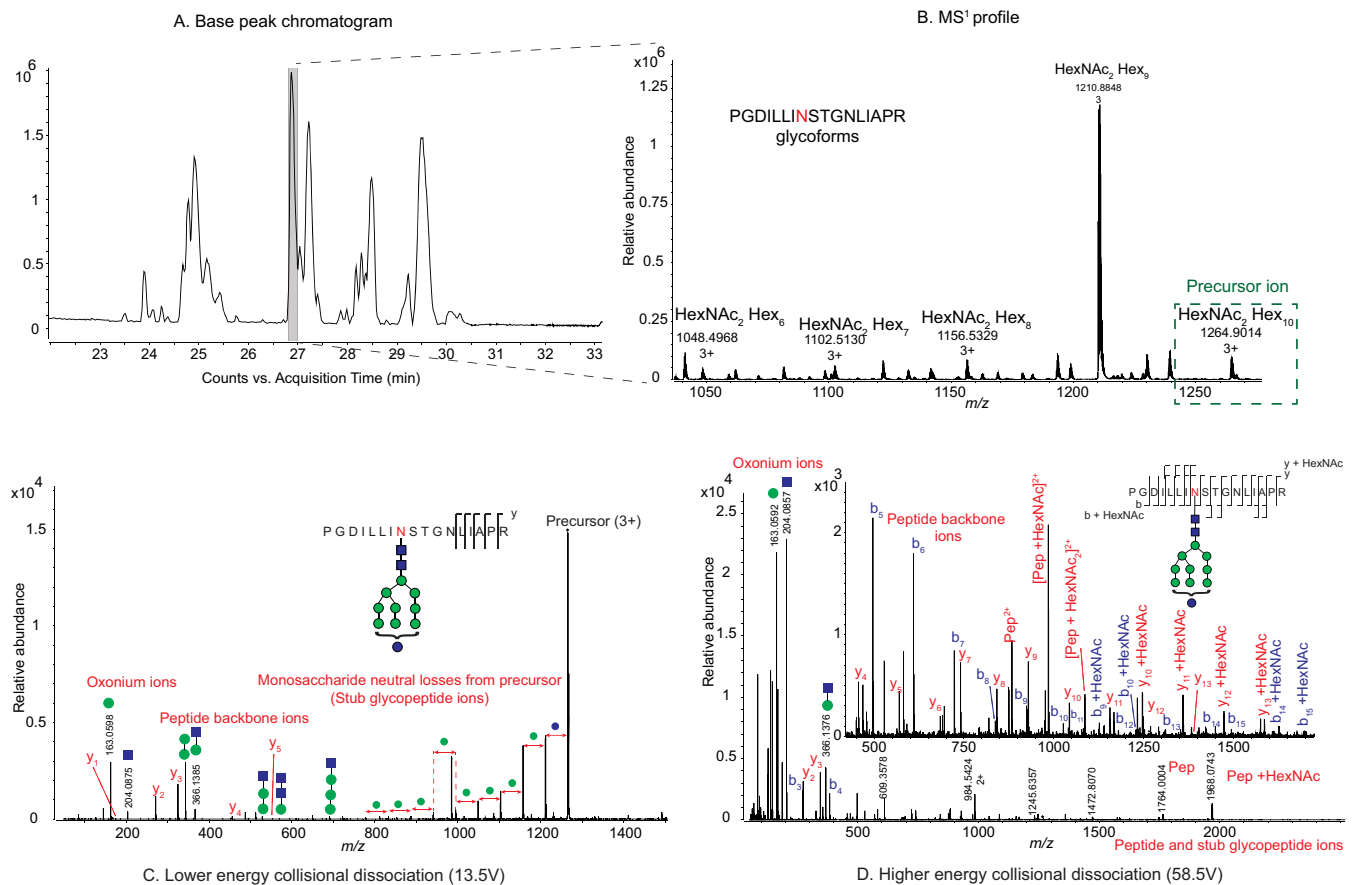
**FIG. 3. Glycomics results from negative mode LC-MS profiling of released IAV *N*-glycans.** A, Comparison of the three strains. Stacked bars represent mean composite relative abundances of different glycan classes, categorized by HexNAc and Hexose units in the identified glycan compositions (Paucimannose: HexNAc = 2, Hexose ≤ 4; High Mannose: HexNAc = 2, Hexose > 5; Hybrid: HexNAc = 3; Bi-antennary: HexNAc = 4; Tri-antennary: HexNAc = 5; Tetra-antennary: HexNAc = 6 and Penta-antennary: HexNAc = 7). Bar plots show individual glycan relative abundances for B, Phil-82; C, Phil-BS; D, PR-08. Colored bars represent mean abundances relative to the most abundant composition detected for each sample. Error bars represent standard deviation. The most abundant glycoforms in each category are labeled with putative topologies using the Consortium for Functional Glycomics glycan representation scheme.

As seen in Fig. 4C, lower-energy collisional dissociation yields abundant ions from glycosidic bond dissociation. Higher collision energies (Fig. 4D) yield tandem mass spectra that contain glycan oxonium ions, peptide backbone dissociation ions and intact-peptide with glycan additions (known as stub-glycopeptides). We used these features in our automated data analysis workflow for glycopeptide assignment.

**Integrated-omics**—Analysis of intact glycopeptides is often challenging because of the complexity and size of the bioinformatics search space (33, 94). Typically, hypotheses for glycopeptide mass spectrometry data analysis are limited by assumptions about the peptide variants and glycoforms

present in a sample. This limitation has been overcome in proteomics analyses with the use of error-tolerant database searches (95–98), which check for modifications (37) and mutations that best explain the data, in addition to those specified by the user. This is difficult to perform in case of glycopeptides because of the already large number of glycoforms associated with each glycosylation site. A number of software tools have emerged recently that mine only glycopeptide tandem MS (99–102) or perform both proteomics and glycoproteomics searches (103). However, the analyses are limited to user defined boundaries of proteoforms and glycoforms. Use of assumptions while defining these boundaries or search-spaces can lead to under or





**FIG. 4. Glycoproteomics analysis data dimensions.** *A*, Chromatographic separation and associated peak areas of enriched Phil-82 IAV glycopeptides; *B*, MS<sup>1</sup> profiling and intact mass assignments of Phil-82 hemagglutinin glycopeptide PGDILLINSTGNLIAPR glycoforms; Tandem MS of glycopeptide PGDILLINSTGNLIAPR - Hex<sub>10</sub>HexNAC<sub>2</sub> showing product ions resulting from lower-energy collisional dissociation (*C*) and higher-energy collisional dissociation (*D*). Glycan topologies shown are speculative based on inferred glycan compositions.

over-estimation of the proteoforms, thus leading to loss of information or high false discovery rates, respectively.

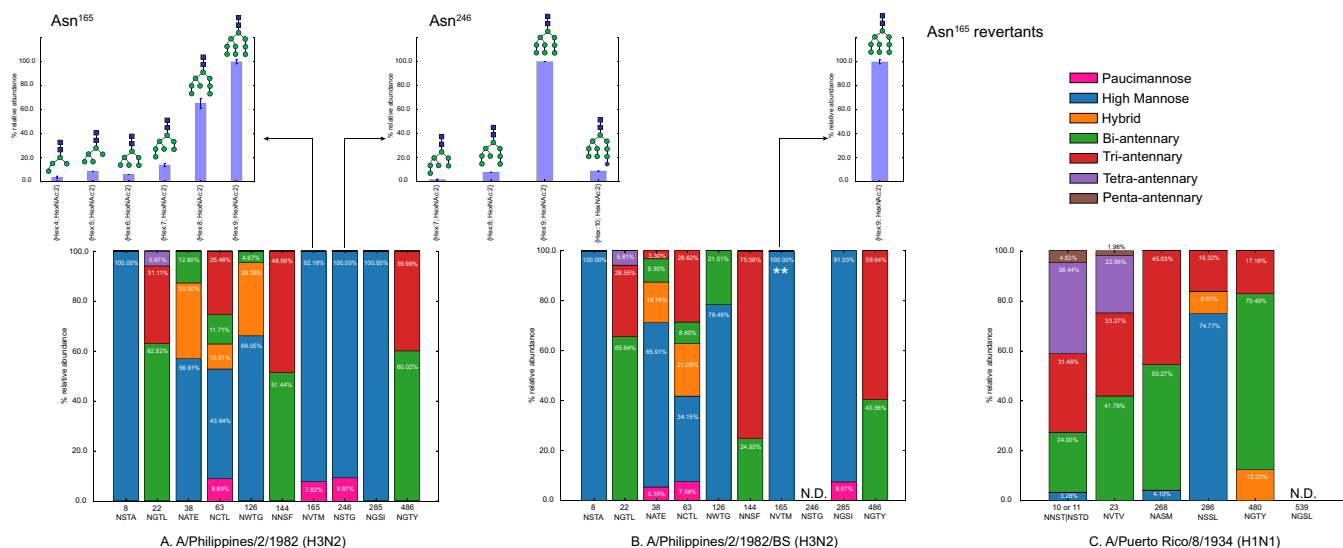
For a glycoprotein as complex as HA, tandem MS can be used to resolve ambiguous assignments; however, glycopeptide tandem MS alone does not allow adequate depth and confidence in assignment for glycoproteoforms. Additionally, glycoproteomics alone does not account for possible mutations incorporated into viral genome as a result of evolutionary pressures.

To produce the most confident information on influenza glycoproteome, we adhered to a *lex parsimoniae* approach, described here as “integrated-omics.” Proteomics and glycomics workflows are well-established and easy to implement. As a first step in the workflow, we performed bottom-up proteomics on PNGaseF deglycosylated influenza virions. The data were mined using hybrid *de novo* and error-tolerant proteomics database searches, which accounted for all possible PTMs, mutations and peptide variants from sample handling. The released IAV N-glycans were analyzed in a separate experiment and searched against glycan databases to define the glycoform search space. Thus, in an effort to minimize assumptions and efficiently mine glycoproteomics

data, the glycomics and proteomics information were integrated to generate a glycopeptide database, against which the glycopeptide MS and tandem MS data were searched. Tandem MS matches were scored based on fragment ion coverage. To calculate a confidence score, a decoy database containing reversed target sequences was used, where the three residue sequon was maintained in the correct order. For example, glycopeptide SVQEIQEIQTFYFTPN(HexNAC)KT-EDTIFLR had a decoy LFITDEN(HexNAC)KTPTFYFFTQIEQ-IEQVSR. Confidence scores (q-values) were calculated as described by Käll *et al.* (46, 104).

The tandem mass spectrometric assignment scores for some glycopeptides, such as those spanning site 126 (NWTG) were lower than the accepted score threshold, because of incomplete coverage and low precursor and product ion abundances. Manual validation helped confirm the assignments and inclusion in the final results. All automated glycopeptide assignment results have been included as comma separated text files in the data repository.

The site-specific glycosylation plots, shown in Fig. 5, represent normalized percent relative abundances of different glycoform classes at each glycosylation site for the hemag-



**FIG. 5. Integrated-omics results.** Site-specific glycan distributions for hemagglutinins. *A*, Phil-82; *B*, Phil-BS; *C*, PR-08. Stacked bars represent composite mean relative abundances for *N*-glycan compositions categorized by number of HexNAc units, as described above. All stacked bars have been scaled to 100% to represent percentages of individual glycan classes among all glycans identified. Insets show individual glycoform relative abundances for sites identified in the interaction with immune-lectins. Error bars represent standard deviation in three measurements of glycoform abundances. N.D.: Not Detected. \*\*Revertant population at site 165 in Phil-BS.

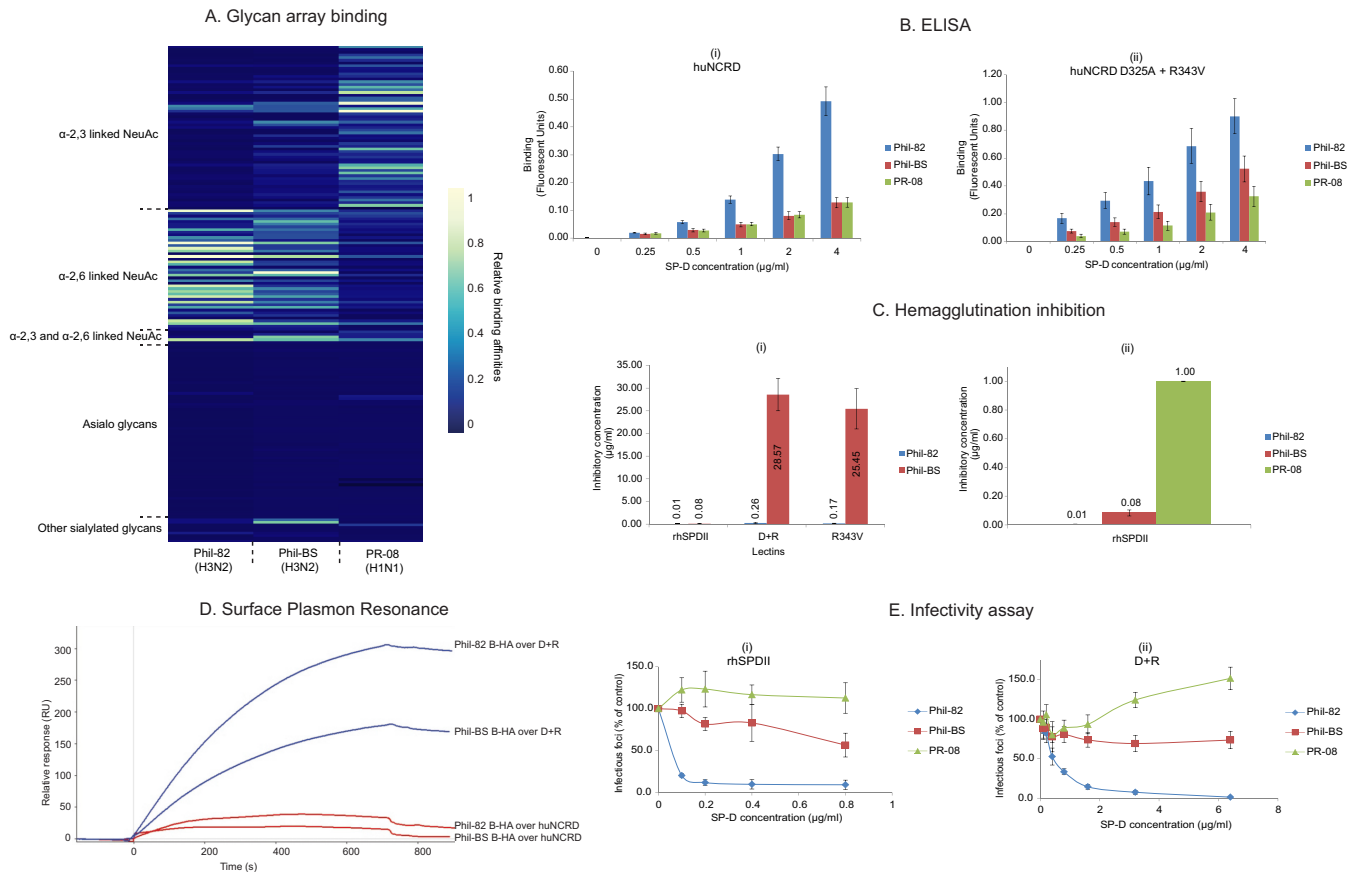
glutinins from three IAV strains. Although the plots present results from glycopeptide LC-tandem MS analyses, the databases against which glycopeptide LC-tandem MS data were searched were generated by combining the proteomics and glycomics results for respective virions. These results therefore reflect the integration of different data domains. Individual glycan compositions contributing to the abundances at sites 165 and 246 have been shown in Fig. 5 (inset). It is clear from the integrated-omics results that glycosylation at Asn<sup>165</sup> and Asn<sup>246</sup> is lost in the escape mutant Phil-BS from the selection pressures inflicted by growth in presence of bovine serum. All site-specific glycan distributions are shown in [supplemental Fig. S3](#), and aligned to individual HA monomer models for better visualization in [supplemental Fig. S4](#). [Supplemental Fig. S5](#) compares glycan distributions from glycomics results with those from integrated-omics results. The figures show good correlation between relative abundances for high-mannose type glycans (2 HexNAc units) and hybrid/complex type glycans (>2 HexNAc units) from glycoproteomics and glycomics data. LC-MS profiling based glycomics is error-prone because multiple glycan compositions can match the given mass and also because noise peaks can sometimes get assigned to glycan compositions. Thus, glycopeptide tandem-MS becomes an essential step in the integrated-omics workflow, to eliminate false-positives and generate confident site-specific glycan profiles.

Hartley *et al.* concluded that Asn<sup>246</sup> on Phil-82 and Phil-BS is not occupied by a glycan (24). These conclusions were derived using results from genetic sequence analyses and mobility shifts in electrophoresis, showing loss of glycosylation in escape mutants. By contrast, we observed that Asn<sup>246</sup>

is in fact glycosylated in Phil-82, whereas Phil-BS loses glycosylation at both Asn<sup>165</sup> and Asn<sup>246</sup>, because of acquired mutations. This exemplifies the need for site-specific structural information on glycoproteins, which genetic sequencing and classical biochemical methods fail to provide.

Consistent with the proteomics results, we identified glycopeptides from a revertant population only for Asn<sup>165</sup> (NVTM); we observed no reversion at Asn<sup>246</sup> (NSTG). To check what percentage of the total virus population reverted to the intact sequon at Asn<sup>165</sup>, we compared arbitrary unit counts from integrated extracted ion chromatograms in the proteomics data. We also calculated site-occupancy for this revertant population as described above. In the proteomics analyses, 2% of the total Phil-BS population contained the intact revertant sequon (NVTM) at Asn<sup>165</sup>, with a glycan site-occupancy of 92%.

Glycosylation changes were limited to loss of glycosylation at Asn<sup>165</sup> and Asn<sup>246</sup>, whereas that at other sequons remained largely unchanged. Both Asn<sup>165</sup> and Asn<sup>246</sup>, present on the HA head, are accessible to host immune lectins. Any differences observed in bioassays could therefore be attributed to differences in glycosylation at these two sites. Although, high-mannose glycans dominated at all sites on H3 HA head, the level of glycan processing varied. In particular, for Phil-82, we noticed a striking difference in extent of high mannose *N*-glycan processing between Asn<sup>165</sup> and Asn<sup>246</sup>. In the early stages of biosynthesis, glycoproteins are occupied by glucosylated high mannose (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>) *N*-glycans. The glucose residues are recognized by calnexin and calreticulin as a quality-control mechanism for protein folding. The three glucose units are removed subsequently in the



**FIG. 6. Results from biochemical, virological and immunological assays.** *A*, CFG glycan array analysis of three IAV strains. Only glycans where relative binding was greater than 0.1% for any one of the three strains have been included in the heat-map. Array glycans are sorted into different categories (y-axis) and the three virus strains are listed on x-axis; *B*, ELISA results. Bar-charts show dose-dependent binding to human surfactant protein-D NCRD wild-type (i) and double mutant D+R (ii). X-axis shows the concentration of SP-D used and Y-axis shows binding response in fluorescent units measured. Error bars represent standard error; *C*, Hemagglutination inhibition of IAV strains. (i) Comparison of inhibitory effect of different lectins on Phil-82 and Phil-BS. (ii) HA inhibitory concentrations of rhSPDII for the three IAV strains studied. Error bars represent standard error; *D*, Surface Plasmon Resonance results. Bromelain cleaved hemagglutinin (B-HA) from Phil-82 and Phil-BS, were introduced in the mobile phase and binding toward immobilized wild type or mutant NCRDs was measured; *E*, Infectivity assay results. Infectivity of the three IAV strains was compared after pre-incubation with increasing concentrations of (i) rhSPDII and (ii) D+R mutant SP-D. Data points show infectivity measured as infectious foci expressing influenza nucleoprotein, relative to control (no SP-D pre-incubation).

endoplasmic reticulum by glucosidases. A significant number of glycoproteins are also acted upon by ER mannosidase-I, which removes a terminal mannose unit, after which the glycoproteins are transported to the Golgi apparatus with a  $\text{Man}_9\text{GlcNAc}_2$  or  $\text{Man}_8\text{GlcNAc}_2$  glycan. In the Golgi apparatus, the mannose residues are trimmed further before becoming extended into complex type structures (105–111). Both  $\text{Asn}^{165}$  and  $\text{Asn}^{246}$  displayed higher abundances of relatively immature glycans containing a higher number of hexose units in their compositions, as seen in the site-specific glycan distributions (Fig. 5 inset and supplemental Fig. S3). The composition  $\text{Hex}_9\text{HexNAc}_2$  was the most abundant at both  $\text{Asn}^{165}$  and  $\text{Asn}^{246}$  and the presence of  $\text{Hex}_{10}\text{HexNAc}_2$  was observed at  $\text{Asn}^{246}$  by tandem MS, as shown in Fig. 4 (C and D).

The glycosylation profile on PR-08 contrasted with those of the two H3N2 strains in that complex type *N*-glycans domi-

nated at all sequons. Given that all three viruses were grown using the same egg culture conditions, this ruled out the involvement of external factors including cell type, growth media and culture system in presence of under-processed *N*-glycans on H3N2 virions. Another recent study has shown similar predominance of high-mannose type glycans on the head region of a different H3N2 strain (27). Although the homologous H3 HAs of the same subtype but different strains show similar glycosylation profiles, the H1 HA from PR-08 contains much more highly processed *N*-glycans. This implicates protein backbone and structural features of the viral subtypes and strains as the drivers in glycan processing and resulting microheterogeneity.

**Bioassays**—We used biochemical and virological assays to survey virion binding preferences to host glycans and correlate differences in glycosylation with susceptibility of HAs toward the immune lectin SP-D (Fig. 6).

**Glycan Array Binding**—In addition to the genetic reassortments that cause antigenic-shift and zoonosis, differences in protein backbone and in glycosylation states affect receptor binding specificities of influenza A virus (87, 112–117). Glycan array results, in Fig. 6A, show binding affinities of the three fluorescently labeled IAV strains to glycans normalized to the strongest binder for each strain. Only glycans where the relative binding was found greater than 0.1% of the strongest bound glycan for any IAV strain have been included. A complete list of glycans and fluorescence readouts from bound IAV are included in Glycan array binding data/results.

Both Phil-82 and Phil-BS strains showed preferential binding to  $\alpha$ -2,6 linked sialic acid containing glycans, versus  $\alpha$ -2,3 linked for the PR-08 H1N1 subtype (Fig. 6A). This showed that the reassorted H3N2 virions, generated with A/Puerto Rico/8/34 backbone, retain the binding specificities on their HAs as expected. Furthermore, the mutations induced by selection pressure leading to loss of two *N*-linked glycans on Phil-BS, did not alter its binding preferences for sialylated glycans and therefore host specificity.

**ELISA**—ELISA results performed at different concentrations of the wild-type huNCRD show stronger binding for Phil-82, compared with Phil-BS and PR-08, which show similar binding affinities, as seen in Fig. 6B(i). The D+R mutant huNCRD has mutations in the amino acids surrounding the lectin site that improve hemagglutination activity (48, 118). Consistent with previous reports, we observed higher binding for all three strains with D+R mutant huNCRD Fig. 6B(ii), the extent of which was substantially higher in the case of Phil-82 and Phil-BS than for PR-08. These results correlated well with our observed abundances for high-mannose type glycans present on the three strains from integrated omics. Although Phil-BS had lower abundance of high-mannose type glycans compared with Phil-82 because of loss of glycosylation at sites 165 and 246, the abundances of immature glycans remained higher than on PR-08, for which mature complex type *N*-glycans predominated.

**Hemagglutination Inhibition**—Surfactant protein-D plays an important role in inhibition of IAV as part of the host innate immune response. SP-D has a carbohydrate recognition domain that binds to mannose containing *N*-glycans on viral hemagglutinin. This presumably inhibits HA binding to host cell surface receptors, preventing docking and entry in to the host cell. Typical hemagglutination inhibition assays are performed using serum to detect the presence and levels of antibodies against HA. Because SP-D prevents IAV interaction with host cells, it can be used to inhibit hemagglutination activity, in a dose-dependent manner (15, 20, 54).

We first determined the sensitivity of Phil-82 and Phil-BS strains toward SP-D lectins. As seen in Fig. 6C(i), rhSPDII (full-length SP-D) was the most effective in inhibiting the hemagglutination activity between the three SP-D constructs compared. Although rhSPDII inhibited both Phil-82 and Phil-BS, the D+R and R343V constructs were effective only

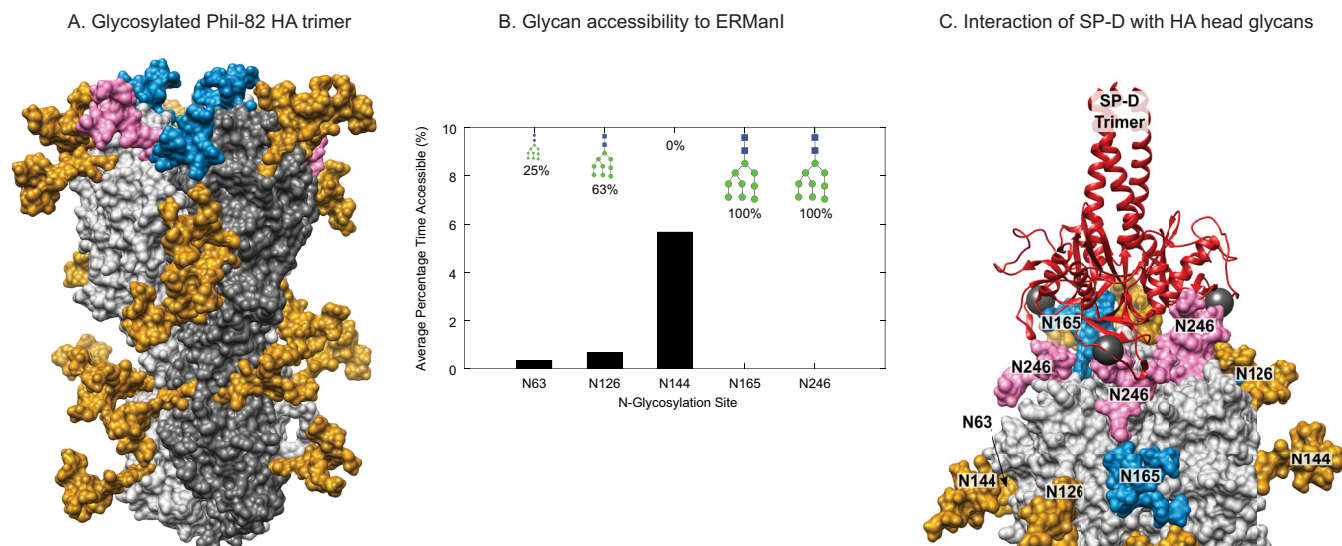
against Phil-82. Fig. 6C(ii) shows the measured inhibitory concentrations of rhSPDII for all three IAV strains. Phil-BS required eightfold higher rhSPDII concentration than Phil-82, whereas PR-08 could not be inhibited even at the highest lectin concentration (1  $\mu$ g/ml), used in the experiment. PR-08 was also found to be insensitive to D+R and R343V (data not shown). We concluded from these ELISA and hemagglutination inhibition assay results that the presence of high-mannose type *N*-glycans on the virus surface correlates with extent of neutralization of hemagglutination by SP-D.

**Surface Plasmon Resonance**—Although HA is almost 10 times more abundant than NA (74), it was important to identify the primary contributor to SP-D interactions and changes therein. We therefore performed SPR experiments using bromelain cleaved hemagglutinins (B-HA) from the wild-type and mutant H3N2 IAV strains. As shown in Fig. 6D, both wild-type and mutant SP-Ds display markedly stronger binding to the Phil-82 B-HA compared with Phil-BS B-HA. Thus, differing SP-D binding among the tested IAV strains can be attributed to changes in the HA. Therefore, ELISA, hemagglutination inhibition and SPR assays established a clear difference in SP-D binding and activity of HA from different IAV strains.

**Infectivity Assay**—We next determined the ability of SP-D to neutralize virus infectivity using fluorescent focus assays. For these experiments we used primary human bronchial/tracheal epithelial cells to ensure that the results reflected behavior of the viruses in non-malignant human respiratory epithelia. Infectivity was measured via production of viral nucleoprotein by infected host cells, using an anti-nucleoprotein antibody. Cells were infected using virions pre-incubated without SP-D or with different concentrations of full-length rhSPDII and the strong binding NCRD mutant D+R. Fig. 6E shows infectivity of the three virus preparations relative to cells infected without SP-D pre-incubation. With increasing concentration of rhSPDII during pre-incubation, the infectivity of Phil-82 declined sharply. By contrast, Phil-BS showed only a slight decrease in infectivity at higher lectin concentrations. With trimeric NCRD D+R, Phil-82 required pre-incubation with higher concentrations of lectin for loss in infectivity whereas no significant loss in infectivity was observed for Phil-BS. PR-08 infectivity was insensitive to either lectin preparation.

These results substantiate the involvement of Asn<sup>165</sup> and Asn<sup>246</sup> in HA interactions with SP-D, a collectin that attenuates viral interactions with host cell surface receptors and causes viral aggregation and recognition by immune cells. We concluded that the presence of high-mannose type glycans at Asn<sup>165</sup> and Asn<sup>246</sup> is crucial toward recognition by collectins.

**Modeling and Molecular Dynamics**—The integrated-omics results and bioassays showed reduced glycan processing at the two sites involved in immune recognition for Phil-82. The predominance of the Hex<sub>9</sub>HexNAc<sub>2</sub> composition at these sites, suggested decreased processing by ERManI. The presence of Hex<sub>10</sub>HexNAc<sub>2</sub> at Asn<sup>246</sup> also suggested a decreased



**FIG. 7. Structure modeling and molecular dynamics results.** *A*, 3D model of glycosylated Phil-82 hemagglutinin trimer. One HA monomer is shown in dark gray. Glycans at Asn<sup>165</sup> and Asn<sup>246</sup> are shown in blue and pink, respectively. *B*, Percentage of simulation time that an *N*-glycan is accessible to ERMAnI. The values are averaged over each protomer of the H3 trimer. The amount of Man<sub>9</sub>GlcNAc<sub>2</sub> at each site is depicted on the upper axis, demonstrating that increased accessibility to ERMAnI correlates with less Man<sub>9</sub>GlcNAc<sub>2</sub> at each site. *C*, A model of trimeric SP-D (red ribbons) bound to three *N*-glycans on the H3 head domain (light gray surface). The 3D structure of the glycosylated H3 was taken from the MD simulation. In this snapshot, Man $\alpha$  residues at sites Asn<sup>246</sup>, Asn<sup>165</sup>, and Asn<sup>246</sup> on different protomers were  $\sim 45$  Å apart, such that SP-D could bind each residue simultaneously. The calcium atoms found in each SP-D Man $\alpha$  binding site are shown (dark gray spheres).

degree of *N*-glycan processing by other enzymes, including glucosidases. We therefore performed three-dimensional (3D) structural modeling and molecular dynamics (MD) simulations to understand the structural basis for the low level of glycan processing at these sites and the interactions of these immature glycans with SP-D.

**Molecular Dynamics Simulation**—A 500 ns simulation was performed on the head domain of the H3 trimer, with Man<sub>9</sub>GlcNAc<sub>2</sub> attached to each of the 15 glycosites (namely Asn<sup>65</sup>, Asn<sup>144</sup>, Asn<sup>126</sup>, Asn<sup>165</sup>, and Asn<sup>246</sup> on each protomer). Despite the initial threefold symmetry of the trimer, unique glycan-glycan and glycan-protein interactions formed on different protomers, a consequence of both the algorithm employed to attach the glycans and of the randomly assigned initial atomic velocities. Over the timescale of the simulation the unique interactions observed on individual protomers were stable and did not interconvert, indicating that interconversion occurs on larger timescales.

**Assessment of Glycan Accessibility to ERMAnI**—The accessibility of each *N*-glycan to ERMAnI was assessed over the course of the simulation. The co-complexed *N*-glycan in the ERMAnI 3D structure was superimposed onto each *N*-glycan on the H3 head domain. Large surface overlaps between ERMAnI and the H3 glycoprotein indicated that the *N*-glycan was inaccessible at that particular time point of the simulation. The accessibility varied depending on the unique structures formed by the glycans on each protomer, but the relative percentage of time an *N*-glycan was accessible to ERMAnI correlated well with the amount of processing observed experimentally (Fig. 7B).

Both Asn<sup>165</sup> and Asn<sup>246</sup> remained completely inaccessible to ERMAnI throughout the simulation, which agrees with the experimental observation that these sites contain only high mannose glycans. Glycans at Asn<sup>63</sup> and Asn<sup>126</sup> which showed some processing to mature glycoforms, become accessible to ERMAnI over the course of the simulation. The most processed site, Asn<sup>144</sup>, adopts shapes accessible to ERMAnI for the longest amount of time during the simulation.

These results demonstrate a clear correlation between the percentage of time an *N*-glycan is accessible to ERMAnI and the degree of processing in the glycoforms observed at a site; however, the results indicate that longer simulation times, not feasible using present state-of-the-art processors, would be required to observe interconversion between the conformations adopted at each site. Further, it should be noted that the actual value of percentage time available is based on an arbitrary overlap cutoff, and that assessment of accessibility does not account for induced fit in ERMAnI or any of the overlapping glycans and protein. Thus it is appropriate to consider only the relative values among sites.

SP-D binding to glycosylated HA - SP-D binds to Man $\alpha$ , which is present in high-mannose *N*-glycans that coat viral surface proteins. As SP-D adopts a trimeric structure with  $\sim 45$  Å between each binding site, we hypothesized that sets of *N*-glycans on the H3 head domain may be positioned such that SP-D could form a multimeric interaction with three *N*-glycans simultaneously, enhancing the interaction strength (see Fig. 7C). It is interesting to note that HAs also have 3-fold symmetry and a  $\sim 45$  Å linear distance between binding sites.

TABLE II

Sets of *N*-glycosylation sites<sup>a</sup> whose spacing and glycan dynamics allow trimeric interaction with a trimeric SP-D. <sup>a</sup>Asn<sup>144</sup> was excluded from the analysis as it was found to have both low site occupancy and a mature glycoform distribution

Sets of N-glycan sites (protomer number)		
Site 1	Site 2	Site 3
Asn <sup>165</sup> (1)	Asn <sup>126</sup> (1)	Asn <sup>63</sup> (3)
	Asn <sup>246</sup> (1)	Asn <sup>126</sup> (3)
Asn <sup>165</sup> (2)	Asn <sup>165</sup> (3)	Asn <sup>63</sup> (1)
	Asn <sup>63</sup> (2)	Asn <sup>126</sup> (3)
Asn <sup>165</sup> (3)	Asn <sup>63</sup> (1)	Asn <sup>126</sup> (2)
Asn <sup>246</sup> (1)	Asn <sup>126</sup> (2)	Asn <sup>126</sup> (3)
	Asn <sup>165</sup> (2)	Asn <sup>126</sup> (3)
		Asn <sup>165</sup> (3)
	Asn <sup>165</sup> (3)	Asn <sup>246</sup> (3)
		Asn <sup>126</sup> (2)
		Asn <sup>126</sup> (3)
	Asn <sup>246</sup> (2)	Asn <sup>126</sup> (1)
	Asn <sup>63</sup> (1)	Asn <sup>126</sup> (2)
Asn <sup>246</sup> (2)	Asn <sup>126</sup> (2)	Asn <sup>165</sup> (3)
		Asn <sup>246</sup> (3)
		Asn <sup>63</sup> (1)
Asn <sup>246</sup> (3)	Asn <sup>126</sup> (2)	Asn <sup>63</sup> (1)
		Asn <sup>63</sup> (2)
Asn <sup>126</sup> (2)	Asn <sup>63</sup> (1)	Asn <sup>63</sup> (2)
	Asn <sup>63</sup> (2)	Asn <sup>126</sup> (3)

We took an ensemble of conformations adopted by the Man<sub>9</sub>GlcNAc<sub>2</sub> glycosylated H3 from the MD simulation. By monitoring the distances between the non-reducing terminal Man<sub>α</sub> residues in each *N*-glycan over the course of the simulation, we determined which sets of three *N*-glycans were positioned for trimeric interaction with SP-D (sets that were 43–47 Å apart). Of the 20 total sets, 18 contained either Asn<sup>165</sup> or Asn<sup>246</sup> (Table II). The data imply that removal of Asn<sup>165</sup> and Asn<sup>246</sup> would greatly reduce the potential for trimeric interactions to form between SP-D and the H3 head domain. This finding agrees with the experimental assays (Fig. 6), where we observed that removal of these two glycosites reduced binding of SP-D.

We observed that the binding sites of trimeric architecture of SP-D are positioned perfectly to interact with glycans on the exposed influenza H3 head domain. Our molecular modeling predicts that the high mannose *N*-glycan at sites Asn<sup>165</sup> and Asn<sup>246</sup> play a pivotal role in forming trimeric interactions with SP-D, in agreement with the experimental evidence.

Enzyme accessibility is only one of the factors that determine the degree of glycan processing at a particular site. Residues outside of the traditional N-X-S/T (where X ≠ P) sequon can profoundly affect both the rate of glycosylation by oligosaccharyltransferase and the predominant glycoform observed (119). Further, mechanisms that govern intracellular proteostasis have been observed to impact the *N*-glycosylation pattern of secreted glycoproteins (120). We note, however, that a clear correlation between the relative accessibility of an *N*-glycan computed from an MD simulation and the

degree of processing has been observed both here and in previous work (111).

**Concluding Remarks**—Several recent studies have underlined the importance of HA glycosylation in influenza virus pathology (5, 114, 121–124). In addition, a number of groups have attempted to trigger a broadly neutralizing antibody response toward influenza hemagglutinins, by artificially modulating HA glycosylation (125, 126). Therefore, HA glycosylation is a clear indicator of virus fitness and evolution (11, 13, 21, 122). Naturally, the ability to predict the fitness of an IAV strain based on its glycosylation profile will help designate strains for vaccine development.

We identified trends in viral evolution with a focus on changes in glycosylation patterns. Our results detailed the site-specific structural changes in HA glycosylation that occur as IAV adapts to immune pressure. We showed that not only the presence or absence of glycosylation, but also the level of glycan processing need to be considered with regard to prediction of interaction with molecules of the host immune system.

Most importantly, we demonstrated how comprehensive analysis of site-specific glycosylation fills in the gaps between modeling and structural studies. We demonstrated that with detailed site-specific glycosylation information, and computational glycosylation modeling, we can identify structural determinants of site-specific glycoforms. Our study describes a new strategy to elucidate IAV-host interactions and controlled evolution studies for understanding the effects of selection pressures in virus evolution.

**Acknowledgments**— We thank Dr. Michael J. Rynkiewicz and Prof. Barbara A. Seaton for helpful discussions; Prof. Erika C. Crouch (Washington University School of Medicine) for providing Surface Plasmon Resonance data. We also thank Thermo-Fisher Scientific for providing access to the Q-Exactive Plus mass spectrometer used in this work.

\* This work was supported by NIH grants P41GM104603 (JZ, KK, JAK), NIH R01 GM100058, and P41 GM103390 (OCG, RJW) and R01 HL069031 (KLH and MRW). D.F. Smith and R.D. Cummings carried out the CFG glycan array analyses at the National Center for Functional Glycomics (Emory University School of Medicine) funded by the NIH grant P41GM103694. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

§ This article contains supplemental material.

\*\* To whom correspondence should be addressed: Boston University Medical Campus, 670 Albany St., Rm. 509, Boston, MA 02118. Tel.: 1–617–638–6762; E-mail: jzaia@bu.edu.

REFERENCES

- Blackburne, B. P., Hay, A. J., and Goldstein, R. A. (2008) Changing selective pressure during antigenic changes in human influenza H3. *PLoS Pathog.* 4, e1000058
- Cherry, J. L., Lipman, D. J., Nikolskaya, A., and Wolf, Y. I. (2009) Evolutionary dynamics of N-glycosylation sites of influenza virus hemagglutinin. *PLoS Curr.* 1, RRN1001
- Vigerust, D. J., and Shepherd, V. L. (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol.* 15, 211–218
- Wei, C. J., Boyington, J. C., Dai, K., Houser, K. V., Pearce, M. B., Kong,

- W. P., Yang, Z. Y., Tumpey, T. M., and Nabel, G. J. (2010) Cross-neutralization of 1918 and 2009 influenza viruses: role of glycans in viral evolution and vaccine design. *Sci. Transl. Med.* **2**, 24ra21–24ra21
5. Magadán, J. G., Altman, M. O., Ince, W. L., Hickman, H. D., Stevens, J., Chevalier, A., Baker, D., Wilson, P. C., Ahmed, R., Bennink, J. R., and Yewdell, J. W. (2014) Biogenesis of influenza A virus hemagglutinin cross-protective stem epitopes. *PLoS Pathog.* **10**, e1004204
  6. He, J. L., Chiu, Y. C., Chang, S. C., Wang, C. H., and Juang, R. H. (2015) Glycosylation at hemagglutinin Asn-167 protects the H6N1 avian influenza virus from tryptic cleavage at Arg-201 and maintains the viral infectivity. *Virus Res.* **197**, 101–107
  7. Carrat, F., and Flahault, A. (2007) Influenza vaccine: The challenge of antigenic drift. *Vaccine* **25**, 6852–6862
  8. Łuksza, M., and Lässig, M. (2014) A predictive fitness model for influenza. *Nature* **507**, 57–61
  9. Cui, J., Smith, T., Robbins, P. W., and Samuelson, J. (2009) Darwinian selection for sites of Asn-linked glycosylation in phylogenetically disparate eukaryotes and viruses. *Proc. Natl. Acad. Sci.* **106**, 13421–13426
  10. Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., and Korber, B. (2010) Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* **14**, 1229–1246
  11. Das, S. R., Puigbò, P., Hensley, S. E., Hurt, D. E., Bennink, J. R., and Yewdell, J. W. (2010) Glycosylation focuses sequence variation in the influenza A virus H1 hemagglutinin globular domain. *PLoS Pathog.* **6**, e1001211
  12. Varki, A. (2006) Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126**, 841–845
  13. Das, S. R., Hensley, S. E., David, A., Schmidt, L., Gibbs, J. S., Puigbò, P., Ince, W. L., Bennink, J. R., and Yewdell, J. W. (2011) Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proc. Natl. Acad. Sci.* **108**, E1417–E1422
  14. Hufford, M. M., Richardson, G., Zhou, H., Manicassamy, B., Garcia-Sastre, A., Enelow, R. I., and Braciale, T. J. (2012) Influenza-infected neutrophils within the infected lungs act as antigen presenting cells for anti-viral CD8<sup>+</sup> T cells. *PLoS ONE* **7**, e46581
  15. Hartshorn, K. L., Crouch, E. C., White, M. R., Eggleston, P., Tauber, A. I., Chang, D., and Sastry, K. (1994) Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. *J. Clin. Invest.* **94**, 311–319
  16. White, M. R., Doss, M., Boland, P., Teclé, T., and Hartshorn, K. L. (2008) Innate immunity to influenza virus: implications for future therapy. *Expert Rev. Clin. Immunol.* **4**, 497–514
  17. Job, E. R., Bottazzi, B., Short, K. R., Deng, Y. M., Mantovani, A., Brooks, A. G., and Reading, P. C. (2014) A single amino acid substitution in the hemagglutinin of H3N2 subtype influenza A viruses is associated with resistance to the long pentraxin PTX3 and enhanced virulence in mice. *J. Immunol.* **192**, 271–281
  18. Qi, L., Kash, J. C., Dugan, V. G., Jagger, B. W., Lau, Y. F., Sheng, Z. M., Crouch, E. C., Hartshorn, K. L., and Taubenberger, J. K. (2011) The ability of pandemic influenza virus hemagglutinins to induce lower respiratory pathology is associated with decreased surfactant protein D binding. *Virology* **412**, 426–434
  19. Hartshorn, K. L., Webby, R., White, M. R., Teclé, T., Pan, C., Boucher, S., Moreland, R. J., Crouch, E. C., and Scheule, R. K. (2008) Role of viral hemagglutinin glycosylation in anti-influenza activities of recombinant surfactant protein D. *Respir. Res.* **9**, 65
  20. Hawgood, S., Brown, C., Edmondson, J., Stumbaugh, A., Allen, L., Goerke, J., Clark, H., and Poulain, F. (2004) Pulmonary collectins modulate strain-specific influenza a virus infection and host responses. *J. Virol.* **78**, 8565–8572
  21. Tate, M. D., Job, E. R., Deng, Y. M., Gunalan, V., Maurer-Stroh, S., and Reading, P. C. (2014) Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* **6**, 1294–1316
  22. Anders, E. M., Hartley, C. A., and Jackson, D. C. (1990) Bovine and mouse serum beta inhibitors of influenza A viruses are mannose-binding lectins. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4485–4489
  23. Hartley, C. A., Jackson, D. C., and Anders, E. M. (1992) Two distinct serum mannose-binding lectins function as beta inhibitors of influenza virus: identification of bovine serum beta inhibitor as conglutinin. *J. Virol.* **66**, 4358–4363
  24. Hartley, C. A., Reading, P. C., Ward, A. C., and Anders, E. M. (1997) Changes in the hemagglutinin molecule of influenza type A (H3N2) virus associated with increased virulence for mice. *Arch. Virol.* **142**, 75–88
  25. Reading, P. C., Pickett, D. L., Tate, M. D., Whitney, P. G., Job, E. R., and Brooks, A. G. (2009) Loss of a single N-linked glycan from the hemagglutinin of influenza virus is associated with resistance to collectins and increased virulence in mice. *Respir. Res.* **10**, 117
  26. Sun, X., Jayaraman, A., Maniprasad, P., Raman, R., Houser, K. V., Pappas, C., Zeng, H., Sasisekharan, R., Katz, J. M., and Tumpey, T. M. (2013) N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. *J. Virol.* **87**, 8756–8766
  27. An, Y., McCullers, J. A., Alymova, I., Parsons, L. M., and Cipollo, J. F. (2015) Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. *J. Proteome Res.* **14**, 3957–3969
  28. Wilson, I. A., Ladner, R. C., Skehel, J. J., and Wiley, D. C. (1983) The structure and role of the carbohydrate moieties of influenza virus haemagglutinin. *Biochem. Soc. Trans.* **11**, 145–147
  29. Skehel, J. J., and Wiley, D. C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569
  30. Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. (2006) Structure and Receptor Specificity of the Hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410
  31. Reading, P. C., Hartley, C. A., Ezekowitz, R. A., and Anders, E. M. (1995) A serum mannose-binding lectin mediates complement-dependent lysis of influenza virus-infected cells. *Biochem. Biophys. Res. Commun.* **217**, 1128–1136
  32. Staples, G. O., Naimy, H., Yin, H., Kileen, K., Kraiczek, K., Costello, C. E., and Zaia, J. (2010) Improved hydrophilic interaction chromatography LC/MS of heparinoids using a chip with postcolumn makeup flow. *Anal. Chem.* **82**, 516–522
  33. Khatri, K., Staples, G. O., Leymarie, N., Leon, D. R., Turiák, L., Huang, Y., Yip, S., Hu, H., Heckendorf, C. F., and Zaia, J. (2014) Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J. Proteome Res.* **13**, 4347–4355
  34. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics MCP* **11**, M111.010587
  35. Magrane, M., and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database J. Biol. Databases Curation* 2011,
  36. Squires, R. B., Noronha, J., Hunt, V., Garcia-Sastre, A., Macken, C., Baumgarth, N., Suarez, D., Pickett, B. E., Zhang, Y., Larsen, C. N., Ramsey, A., Zhou, L., Zaremba, S., Kumar, S., Deitrich, J., Klem, E., and Scheuermann, R. H. (2012) Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir. Viruses* **6**, 404–416
  37. Creasy, D. M., and Cottrell, J. S. (2004) Unimod: Protein modifications for mass spectrometry. *PROTEOMICS* **4**, 1534–1536
  38. Horn, D. M., Zubarev, R. A., and McLafferty, F. W. (2000) Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **11**, 320–332
  39. Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J. N., Anderson, G. A., and Smith, R. D. (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* **10**, 87
  40. Ranzinger, R., Herget, S., Wetter, T., and von der Lieth, C. W. (2008) GlycomeDB - integration of open-access carbohydrate structure databases. *BMC Bioinformatics* **9**, 384
  41. Ranzinger, R., Herget, S., von der Lieth, C. W., and Frank, M. (2011) GlycomeDB—a unified database for carbohydrate structures. *Nucleic Acids Res.* **39**, D373–D376
  42. Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., Conley, S., Staples, G. O., Slysz, G. W., Smith, R. D., and Zaia, J. (2012) GlycReSoft: a software package for automated recognition of glycans from LC/MS data. *PLoS ONE* **7**, e45474
  43. Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W. H., Römpf, A., Neumann, S., Pizarro, A. D., Montec-

- chi-Palazzi, L., Tasman, N., Coleman, M., Reisinger, F., Souda, P., Hermjakob, H., Binz, P. A., and Deutsch, E. W. (2011) mzML—a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133
44. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536
45. Kaur, P., and O'Connor, P. B. (2006) Algorithms for automatic interpretation of high resolution mass spectra. *J. Am. Soc. Mass Spectrom.* **17**, 459–468
46. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7**, 29–34
47. Tripathi, S., White, M. R., and Hartshorn, K. L. (2015) The amazing innate immune response to influenza A virus infection. *Innate Immun.* **21**, 73–98
48. Crouch, E., Hartshorn, K., Horlacher, T., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Seeberger, P. H., and Head, J. (2009) Recognition of mannosylated ligands and influenza A virus by human surfactant protein D: contributions of an extended site and residue 343†,‡. *Biochemistry* **48**, 3335–3345
49. Nikolaidis, N. M., White, M. R., Allen, K., Tripathi, S., Qi, L., McDonald, B., Taubenberger, J., Seaton, B. A., McCormack, F. X., Crouch, E. C., and Hartshorn, K. L. (2014) Mutations flanking the carbohydrate binding site of surfactant protein D confer antiviral activity for pandemic influenza A viruses. *Am. J. Physiol.* **306**, L1036–L1044
50. Crouch, E., McDonald, B., Smith, K., Cafarella, T., Seaton, B., and Head, J. (2006) Contributions of phenylalanine 335 to ligand recognition by human surfactant protein D ring interactions with SP-D ligands. *J. Biol. Chem.* **281**, 18008–18014
51. Hartshorn, K. L., White, M. R., Smith, K., Sorensen, G., Kuroki, Y., Holmskov, U., Head, J., and Crouch, E. C. (2010) Increasing antiviral activity of surfactant protein D trimers by introducing residues from bovine serum collectins: dissociation of mannan-binding and antiviral activity. *Scand. J. Immunol.* **72**, 22–30
52. Crouch, E., McDonald, B., Smith, K., Roberts, M., Mealy, T., Seaton, B., and Head, J. (2007) Critical role of Arg/Lys343 in the species-dependent recognition of phosphatidylinositol by pulmonary surfactant protein D. *Biochemistry* **46**, 5160–5169
53. Crouch, E., Nikolaidis, N., McCormack, F. X., McDonald, B., Allen, K., Rynkiewicz, M. J., Cafarella, T. M., White, M., Lewnard, K., Leymarie, N., Zaia, J., Seaton, B. A., and Hartshorn, K. L. (2011) Mutagenesis of surfactant protein D informed by evolution and x-ray crystallography enhances defenses against influenza A virus in vivo. *J. Biol. Chem.* **286**, 40681–40692
54. Goh, B. C., Rynkiewicz, M. J., Cafarella, T. R., White, M. R., Hartshorn, K. L., Allen, K., Crouch, E. C., Calin, O., Seeberger, P. H., Schulten, K., and Seaton, B. A. (2013) Molecular mechanisms of inhibition of influenza by surfactant protein D revealed by large-scale molecular dynamics simulation. *Biochemistry* **52**, 8527–8538
55. Hartshorn, K. L., White, M. R., Shepherd, V., Reid, K., Jensenius, J. C., and Crouch, E. C. (1997) Mechanisms of anti-influenza activity of surfactant proteins A and D: comparison with serum collectins. *Am. J. Physiol.* **273**, L1156–L1166
56. White, M. R., Kandel, R., Tripathi, S., Condon, D., Qi, L., Taubenberger, J., and Hartshorn, K. L. (2014) Alzheimer's associated  $\beta$ -amyloid protein inhibits influenza A virus and modulates viral interactions with phagocytes. *PLoS ONE* **9**, e101364
57. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C. H., and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17033–17038
58. Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S., and Sasisekharan, R. (2006) Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology* **16**, 82R–90R
59. Smith, D. F., Song, X., and Cummings, R. D. (2010) Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol.* **480**, 417–444
60. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins Struct. Funct. Bioinforma.* **23**, 318–326
61. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006) Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis A10* **5**, Unit-5.6
62. Fiser, A., and Šali, A. (2003) in *Macromolecular Crystallography, Part D*, ed Enzymology B-M in (Academic Press), pp 461–491
63. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655
64. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103–114
65. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., and Woods, R. J. (2014) Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **35**, 526–539
66. Case, D., Berryman, J., Betz, R., Cerutti, D., Cheatham III, T., Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S. LeGrand, S. Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R. Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York D. M., and Kollman, P. A. (2015) AMBER 2015. *Univ. Calif. San Franc.*, D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman (2015), AMBER 2015, University of California, San Francisco.
67. Mahoney, M. W., and Jorgensen, W. L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910–8922
68. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555
69. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888
70. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
71. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341
72. Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095
73. Eijk, M. van, Rynkiewicz, M. J., White, M. R., Hartshorn, K. L., Zou, X., Schulten, K., Luo, D., Crouch, E. C., Cafarella, T. R., Head, J. F., Haagsman, H. P., and Seaton, B. A. (2012) A unique sugar-binding site mediates the distinct anti-influenza activity of pig surfactant protein D. *J. Biol. Chem.* **287**, 26666–26677
74. Hutchinson, E. C., Charles, P. D., Hester, S. S., Thomas, B., Trudgian, D., Martínez-Alonso, M., and Fodor, E. (2014) Conserved and host-specific features of influenza virion architecture. *Nat. Commun.* **5**, 4816
75. An, H. J., Froehlich, J. W., and Lebrilla, C. B. (2009) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.* **13**, 421–426
76. Desaire, H. (2013) Glycopeptide analysis: recent developments and applications. *Mol. Cell. Proteomics*, **12**, 893–901



77. Hong, Q., Lebrilla, C. B., Miyamoto, S., and Ruhaak, L. R. (2013) Absolute quantitation of immunoglobulin G and its glycoforms using multiple reaction monitoring. *Anal. Chem.* **85**, 8585–8593
78. Thaysen-Andersen, M., and Packer, N. H. (2014) Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta* **1844**, 1437–1452
79. Woodin, C. L., Maxon, M., and Desaire, H. (2013) Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *The Analyst* **138**, 2793
80. Dallas, D. C., Martin, W. F., Hua, S., and German, J. B. (2013) Automated glycopeptide analysis—review of current state and future directions. *Brief. Bioinform.* **14**, 361–374
81. Couceiro, J. N., Paulson, J. C., and Baum, L. G. (1993) Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res.* **29**, 155–165
82. Suzuki, Y., Ito, T., Suzuki, T., Holland, R. E., Chambers, T. M., Kiso, M., Ishida, H., and Kawaoka, Y. (2000) Sialic acid species as a determinant of the host range of influenza A viruses. *J. Virol.* **74**, 11825–11831
83. Stevens, J., Blixt, O., Paulson, J. C., and Wilson, I. A. (2006) Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **4**, 857–864
84. Stevens, J., Blixt, O., Chen, L. M., Donis, R. O., Paulson, J. C., and Wilson, I. A. (2008) Recent avian H5N1 viruses exhibit increased propensity for acquiring human receptor specificity. *J. Mol. Biol.* **381**, 1382–1394
85. Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A., and Wiley, D. C. (1983) Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78
86. Glaser, L., Stevens, J., Zamarin, D., Wilson, I. A., García-Sastre, A., Tumpey, T. M., Basler, C. F., Taubenberger, J. K., and Palese, P. (2005) A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *J. Virol.* **79**, 11533–11536
87. Tharakaraman, K., Raman, R., Viswanathan, K., Stebbins, N. W., Jayaraman, A., Krishnan, A., Sasisekharan, V., and Sasisekharan, R. (2013) Structural determinants for naturally evolving H5N1 hemagglutinin to switch its receptor specificity. *Cell* **153**, 1475–1485
88. Bhatia, V. N., Perlman, D. H., Costello, C. E., McComb, M. E. “Software tool for researching annotations of proteins: open-source protein annotation software with data visualization” *Anal. Chem.* **2009**, *81*, 9819–23. PMC2787672.
89. Gupta, R., Jung, E., and Brunak, S. (2004) Prediction of N-glycosylation sites in human proteins. *Prep.*
90. Beigneux, A. P., Gin, P., Davies, B. S., Weinstein, M. M., Bensadoun, A., Ryan, R. O., Fong, L. G., and Young, S. G. (2008) Glycosylation of Asn-76 in mouse GPIHBP1 is critical for its appearance on the cell surface and the binding of chylomicrons and lipoprotein lipase. *J. Lipid Res.* **49**, 1312–1321
91. Vitorino, R., Alves, R., Barros, A., Caseiro, A., Ferreira, R., Lobo, M. C., Bastos, A., Duarte, J., Carvalho, D., Santos, L. L., and Amado, F. L. (2010) Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics* **10**, 3732–3742
92. Job, E. R., Deng, Y. M., Tate, M. D., Bottazzi, B., Crouch, E. C., Dean, M. M., Mantovani, A., Brooks, A. G., and Reading, P. C. (2010) Pandemic H1N1 influenza A viruses are resistant to the antiviral activities of innate immune proteins of the collectin and pentraxin superfamilies. *J. Immunol.* **185**, 4284–4291
93. Gomez, A., Nekvindova, J., Travica, S., Lee, M. Y., Johansson, I., Edler, D., Mkrtchian, S., and Ingelman-Sundberg, M. (2010) Colorectal cancer-specific cytochrome P450 2W1: intracellular localization, glycosylation, and catalytic activity. *Mol. Pharmacol.* **78**, 1004–1011
94. Desaire, H., and Hua, D. (2009) When can glycopeptides be assigned based solely on high-resolution mass spectrometry data? *Int. J. Mass Spectrom.* **287**, 21–26
95. Creasy, D. M., and Cottrell, J. S. (2002) Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2**, 1426–1434
96. Liska, A. J., and Shevchenko, A. (2003) Expanding the organismal scope of proteomics: Cross-species protein identification by mass spectrometry and its implications. *Proteomics* **3**, 19–28
97. Sunyaev, S., Liska, A. J., Golod, A., Shevchenko, A., and Shevchenko, A. (2003) MultiTag: multiple error-tolerant sequence Tag search for the sequence-similarity identification of proteins by mass spectrometry. *Anal. Chem.* **75**, 1307–1315
98. Nesvizhskii, A. I. (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **73**, 2092–2123
99. Ren, J. M., Rejtar, T., Li, L., and Karger, B. L. (2007) N-glycan structure annotation of glycopeptides using a linearized glycan structure database (GlyDB). *J. Proteome Res.* **6**, 3162–3173
100. Goldberg, D., Bern, M., Parry, S., Sutton-Smith, M., Panico, M., Morris, H. R., and Dell, A. (2007) Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J. Proteome Res.* **6**, 3995–4005
101. Joenväärä, S., Ritamo, I., Peltoniemi, H., and Renkonen, R. (2008) N-Glycoproteomics – An automated workflow approach. *Glycobiology* **18**, 339–349
102. Pompach, P., Chandler, K. B., Lan, R., Edwards, N., and Goldman, R. (2012) Semi-automated identification of N-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC–MS/MS, and glycan database search. *J. Proteome Res.* **11**, 1728–1740
103. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis AI Chapter 13, Unit13.20*
104. Zhu, Z., Su, X., Go, E. P., and Desaire, H. (2014) New glycoproteomics software, GlycoPep evaluator, generates decoy glycopeptides de novo and enables accurate false discovery rate analysis for small data sets. *Anal. Chem.* **86**, 9212–9219
105. Jakob, C. A., Burda, P., Roth, J., and Aebi, M. (1998) Degradation of misfolded endoplasmic reticulum glycoproteins in *Saccharomyces cerevisiae* is determined by a specific oligosaccharide structure. *J. Cell Biol.* **142**, 1223–1233
106. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049
107. Wujek, P., Kida, E., Walus, M., Wisniewski, K. E., and Golabek, A. A. (2004) N-glycosylation is crucial for folding, trafficking, and stability of human tripeptidyl-peptidase I. *J. Biol. Chem.* **279**, 12827–12839
108. Mast, S. W., and Moremen, K. W. (2006) Family 47 alpha-mannosidases in N-glycan processing. *Methods Enzymol.* **415**, 31–46
109. Stanley, P., Schachter, H., and Taniguchi, N. (2009) in *Essentials of Glycobiology*, eds Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etlzler, M. E. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 2nd Ed.
110. Aebi, M. (2013) N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1833**, 2430–2437
111. Hang, I., Lin, C., Grant, O. C., Fleurkens, S., Villiger, T. K., Soos, M., Morbidelli, M., Woods, R. J., Gauss, R., and Aebi, M. (2015) Analysis of site-specific N-glycan remodelling in the ER and the Golgi. *Glycobiology*, cww058
112. Treanor, J. (2004) Influenza vaccine — Outmaneuvering antigenic shift and drift. *N. Engl. J. Med.* **350**, 218–220
113. de Vries, R. P., de Vries, E., Bosch, B. J., de Groot, R. J., Rottier, P. J., and de Haan, C. A. (2010) The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology* **403**, 17–25
114. Jayaraman, A., Koh, X., Li, J., Raman, R., Viswanathan, K., Shriver, Z., and Sasisekharan, R. (2012) Glycosylation at Asn<sup>91</sup> of H1N1 haemagglutinin affects binding to glycan receptors. *Biochem. J.* **444**, 429–435
115. Tharakaraman, K., Raman, R., Stebbins, N. W., Viswanathan, K., Sasisekharan, V., and Sasisekharan, R. (2013) Antigenically intact hemagglutinin in circulating avian and swine influenza viruses and potential for H3N2 pandemic. *Sci. Rep.* **3**, 1822
116. Steinhauer, D. A. (2013) Influenza: Pathways to human adaptation. *Nature* **499**, 412–413
117. Xu, R., Vries, R. P. de, Zhu, X., Nycholat, C. M., McBride, R., Yu, W., Paulson, J. C., and Wilson, I. A. (2013) Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. *Science* **342**, 1230–1235
118. Crouch, E., Tu, Y., Briner, D., McDonald, B., Smith, K., Holmskov, U., and Hartshorn, K. (2005) Ligand specificity of human surfactant protein D expression of a mutant trimeric collectin that shows enhanced interactions with influenza A virus. *J. Biol. Chem.* **280**, 17046–17056
119. Murray, A. N., Chen, W., Antonopoulos, A., Hanson, S. R., Wiseman, R. L.,

- Dell, A., Haslam, S. M., Powers, D. L., Powers, E. T., and Kelly, J. W. (2015) Enhanced aromatic sequons increase oligosaccharyltransferase glycosylation efficiency and glycan homogeneity. *Chem. Biol.* **22**, 1052–1062
120. Dewal, M. B., DiChiara, A. S., Antonopoulos, A., Taylor, R. J., Harmon, C. J., Haslam, S. M., Dell, A., and Shoulders, M. D. (2015) XBP1s links the unfolded protein response to the molecular architecture of mature N-glycans. *Chem. Biol.* **22**, 1301–1312
121. Carbone, V., Schneider, E. K., Rockman, S., Baker, M., Huang, J. X., Ong, C., Cooper, M. A., Yuriev, E., Li, J., and Velkov, T. (2015) Molecular characterisation of the haemagglutinin glycan-binding specificity of egg-adapted vaccine strains of the pandemic 2009 H1N1 swine influenza A virus. *Molecules* **20**, 10415–10434
122. Medina, R. A., Stertz, S., Manicassamy, B., Zimmermann, P., Sun, X., Albrecht, R. A., Uusi-Kerttula, H., Zagordi, O., Belshe, R. B., Frey, S. E., Tumpey, T. M., and García-Sastre, A. (2013) Glycosylations in the globular head of the hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza viruses. *Sci. Transl. Med.* **5**, 187ra70–187ra70
123. Hütter, J., Rödig, J. V., Höper, D., Seeberger, P. H., Reichl, U., Rapp, E., and Lepenies, B. (2013) Toward animal cell culture-based influenza vaccine design: viral hemagglutinin N-glycosylation markedly impacts immunogenicity. *J. Immunol.* **190**, 220–230
124. Zhang, Y., Zhu, J., Li, Y., Bradley, K. C., Cao, J., Chen, H., Jin, M., and Zhou, H. (2013) Glycosylation on hemagglutinin affects the virulence and pathogenicity of pandemic H1N1/2009 influenza A virus in mice. *PLoS ONE* **8**, e61397
125. Lu, Y., Welsh, J. P., and Swartz, J. R. (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proc. Natl. Acad. Sci.* **111**, 125–130
126. Chen, J. R., Yu, Y. H., Tseng, Y. C., Chiang, W. L., Chiang, M. F., Ko, Y. A., Chiu, Y. K., Ma, H. H., Wu, C. Y., Jan, J. T., Lin, K. I., Ma, C., and Wong, C. H. (2014) Vaccination of monoglycosylated hemagglutinin induces cross-strain protection against influenza virus infections. *Proc. Natl. Acad. Sci.* **111**, 2476–2481